# Regulation of gastrulation through dynamic patterning in the *Drosophila* embryo

Thesis by

Nathanie Alna Trisnadi

In Partial Fulfillment of the Requirements for the degree
of
Doctor of Philosophy


CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2015

(Defended 9 July 2014)

# ACKNOWLEDGEMENTS

Many thanks and praises go to my advisor, Angela Stathopoulos. She has been invaluable in my growth as a scientist. I admire her exuberance and imagination that inspire thought-provoking ideas. It has been a joy to work in her lab, and I appreciate her constant guidance as I navigated my graduate studies. Angela has set a high bar for future mentors!

Special thanks to my committee – David Chan, Scott Fraser, Bruce Hay, and Kai Zinn – for their support, feedback, and time. April 1st has been a wonderful day to schedule meetings with this brilliant group. No joke.

Thanks to my collaborators for their endless patience and energy throughout the multiple rounds of manuscript editing and Skype conferences: Gregory T. Reeves, Thai V. Truong, Young-Kyung Bae, Alphan Altinok, Marcos Nahmad, Sophie Katz, and Snehalata Kadam. Likewise, the Zinn lab has been gracious in providing me with *Drosophila* lines, especially when I inadvertently killed stocks and would sheepishly request another vial… or two… ok, maybe more than three. I also appreciate everyone at Caltech's Biological Imaging Center. You all have taught me a lot about microscopes, image analysis, and sitting in front of a computer inside a cold, dark room.

To my fellow labmates, both past and present, thank you for all your help and humor that have made my experiments possible and my time at Caltech exceptionally fun.

And to my dear friends and family, I am grateful for your love throughout this journey.

*From Him the whole body, joined and knit together by every supporting ligament, grows and builds itself up in love, as each part does its work. – Ephesians 4:16*

# ABSTRACT

Gene patterning delineates an embryo into precise domains of differential gene expression. However, throughout gastrulation, these patterns are spatiotemporally dynamic due to the changing environment inherent in development and the contribution of multiple inputs. We investigated how the spatiotemporal dynamics of gene expressions influence two processes in the early *Drosophila* embryo: the establishment of the dorsal-ventral axis and the subsequent mesoderm migration. We found that genes are able to integrate many forms of regulation over space and time in order to refine their expression boundaries and guide gastrulation. Live imaging of the Dorsal transcription factor morphogen gradient revealed spatiotemporal dynamics that never reached steady state. Computational simulations correlated these changes with shifts in the boundaries of downstream target genes. For early mesoderm development, we conducted a screen to ectopically express proteins in specific domains to identify factors involved in migration. We showed that modulation of fibroblast growth factor (FGF) signaling switches between two proteoglycans to transition cells from migration to differentiation. In addition, multiple contributions regulate the complementary expression of cadherins, which is required to provide the proper balance of cell-cell interactions during mesoderm migration. We conclude that the changing environment of the embryo is an important factor during gastrulation and give examples of its impact in defining gene expression domains, supporting specificity of signaling pathways, and regulating adhesion during collective movements.

# TABLE OF CONTENTS

# ABBREVIATIONS

| | |
|---|---|
| AP | anterior-posterior |
| CSPG | chondroitin sulfate proteoglycan |
| CVM | caudal visceral mesoderm |
| Dl | Dorsal |
| DV | dorsal-ventral |
| Ecad | E-cadherin (*Drosophila* Shotgun, Shg) |
| ECM | extracellular matrix |
| EMT | epithelial-to-mesenchymal transition |
| FGF | fibroblast growth factor |
| HSPG | heparan sulfate proteoglycan |
| Htl | Heartless (FGF receptor) |
| MBT | mid-blastula transition |
| MZT | maternal-to-zygotic transition |
| nc | nuclear cycle |
| Ncad | N-cadherin |
| Pyr | Pyramus (FGF ligand) |
| Sdc | Syndecan (HSPG) |
| Sna | Snail |
| Ths | Thisbe (FGF ligand) |
| Trol | Terribly reduced optic lobes (HSPG) |
| 2P-SPIM | two-photon scanned light-sheet (selective plane illumination) microscopy |

CHAPTER I

# Introduction

**Spatiotemporal dynamics of gene expression affect the early *Drosophila* embryo**

There are many different levels of regulation to consider when studying the developing embryo. Gene regulatory networks have mapped out the interactions contributing to gene expression. However the environment surrounding the cell is constantly changing, and it is in this backdrop that an embryo has to develop. We discuss the spatiotemporal dynamics of gene expressions, which occur on the timescale of minutes in the early *Drosophila* embryo, and consider their influence on patterning the embryo.

Initial patterning transpires when successive nuclear divisions terminate molecular processes, including transcription. Therefore, early gene expressions arise from maternal products. As these nuclear cycles decelerate, zygotic transcription escalates and increases the number of inputs contributing to spatially divide the embryo for future differentiation.

*Temporal constraints in gene patterning*

The first three hours of the *Drosophila* embryonic development are comprised of thirteen division events within fourteen nuclear cycles. This occurs in a syncytium where there are no cellular membranes. The first nine nuclear cycles are sequential rounds of synthesis and mitosis encompassing roughly eight minutes at 25°C (Lee and Orr-Weaver, 2003). There is no gap phase in these cycles and multiple origins of replication allow a fast replication time of less than four minutes. At nuclear cycle (nc) 10, DNA synthesis becomes progressively slower as a consequence of fewer replication origins until it reaches 14 minutes at nc 13 (Shermoen et al., 2010). It is also during nuclear cycles 9 and 10 that the nuclei, which are initially dispersed throughout the embryo, migrate from the yolk to

spread evenly along the periphery of the embryo (Table 1, Figure 1). The regulation of these nuclear cycles, which in turn impact gene expression, has been extensively studied.

| nc | Length (min) | Number of Nuclei | |
|---|---|---|---|
| 1 | 8.6 | 1 | Multiple origins of replication; synthesis finishes in 3.4 minutes. |
| 2 | 8.6 | 2 | |
| 3 | 8.6 | 4 | |
| 4 | 8.6 | 8 | |
| 5 | 8.6 | 16 | |
| 6 | 8.6 | 32 | |
| 7 | 8.6 | 64 | Earliest reported zygotic transcription. |
| 8 | 8.6 | 128 | |
| 9 | 8.6 | 256 | |
| 10 | 10.3 | 512 | Synthesis slows down due to less replication origins; mitosis length remains the same. Nuclei complete migration to periphery of embryo. |
| 11 | 11.4 | (750) | *(estimated number of nuclei based on surface density)* |
| 12 | 14 | (1500) | |
| 13 | 19.3 | (3000) | Maternal-to-zygotic transition (MTZ) begins and continues in nc 14. |
| 14 | 50* | (6000) | *Length of nc 14 taken from synthesis time only. G2 phase is introduced (mid-blastula transition, MBT). The G2 length varies depending on cell type and thus the 14$^{th}$ division begins asynchronously throughout the embryo. The embryo is cellularized during this time. |

**Table 1. Timing of nuclear cycles and major events in the early embryo.** Nuclear cycles (nc) gets progressively longer as the embryo develops. Data from Foe and Alberts, 1983; Shermoen et al., 2010.

Many previous observations pointed to regulation of the timing of these nuclear cycles by a cyclin oscillator model, since their levels accumulate during interphase, which in the case of the *Drosophila* syncytium consists only of the synthesis phase. It was proposed that a cyclin threshold triggers mitotic entry when it is promptly degraded to reset

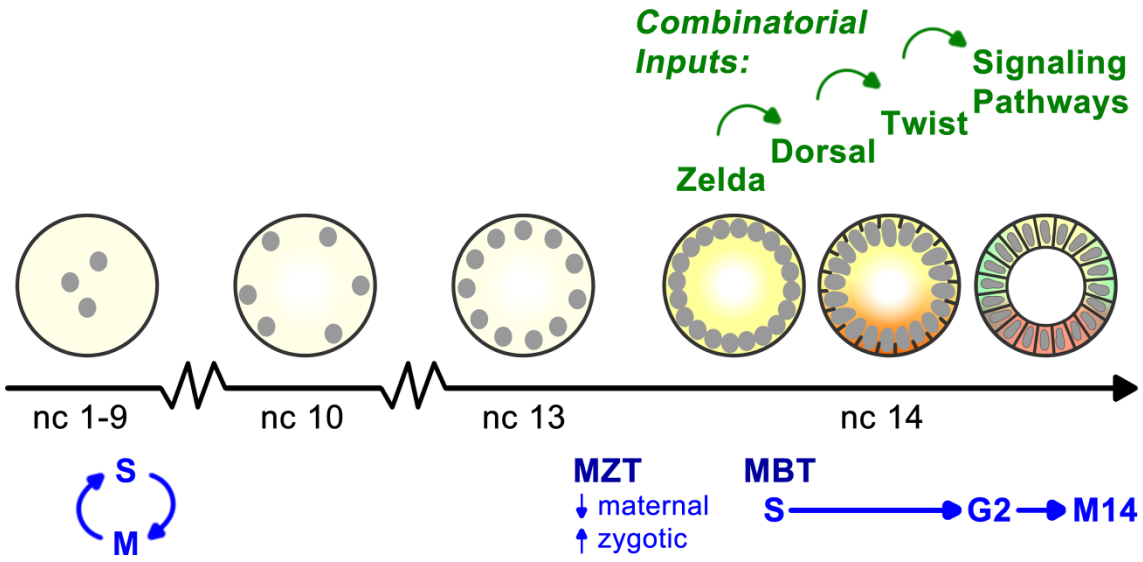



**Figure 1. Summary of events in the *Drosophila* embryo.** Schematic of embryo cross-sections throughout nuclear cycles (nc) 1-14. Nuclear cycles 1-9 are sequential rounds of DNA synthesis (S) and mitosis (M). During nuclear cycles 9 and 10, nuclei (gray circles) migrate to the periphery of the embryo and synthesis gets progressively longer. Throughout nuclear cycles 13 and 14, there is a maternal-to-zygotic transition (MZT) in transcription. A gap phase (G2) is introduced in nc 14, signifying the mid-blastula transition (MBT). During this pause, the embryo is cellularized and gene expressions are refined to their final domains through multiple inputs of transcription factors and signaling pathways (green arrows).

for the next cycle (Edgar et al., 1994; Murray and Kirschner, 1989; O'Farrell, 2001). Cyclins are required for Cyclin-dependent kinases (Cdks) which in turn initiate steps along the cell cycle. However, a recent study decreasing various cyclins using RNAi did not see a proportional lengthening of interphase as expected in order to reach threshold levels (McCleland et al., 2009). These mutants did have defects during mitosis, suggesting an indirect role for cyclins. While cyclins may still schedule cell cycles during early development in other animals (Fung and Poon, 2005), it remains unclear what regulates the timing for *Drosophila* nuclear cycles. Multiple regulatory elements are likely supporting these early, robust division events. Nevertheless, the rapid nuclear cycles are an important

consideration when investigating gene patterning in the embryo. Many studies are now including multiple nuclear cycle stages and dividing longer cycles, such as nc 13 and nc 14, into separate time points (Lott et al., 2011; Reeves et al., 2012). This allows us to detect refinement of gene expression domains, which can occur within a single nuclear cycle.

Mitosis forces transcription factors to disengage from their DNA binding sites, making it difficult to generate new, complete transcripts during the early cycles. Therefore, the limited time during which transcription can occur is constrained by the length of these cycles. *Drosophila* has a reported transcription rate of 1.1-1.4 kilobases (kb) per minute (Shermoen and O'Farrell, 1991; Thummel et al., 1990). The longest gene that could theoretically be transcribed, with a maximum 1.4 kb/min rate and an eight minute nuclear cycle disregarding mitosis, would be 11,200 base pairs. Many, though not all, of the early zygotically transcribed genes are smaller than 10 kb (Reeves et al., 2012). The earliest reported gene that is transcribed, Engrailed, is less than 3 kb long and begins at nc 7 (Ali-Murthy et al., 2013). Yet transcripts do not necessarily accumulate over nuclear cycles like cyclins and instead are degraded at each mitosis, as shown for the homeobox gene *ultrabithorax (ubx)* (Shermoen and O'Farrell, 1991). Aborted transcripts will not be translated and thus pose the question of what function could longer, incomplete mRNA fulfill.

It is possible that gene length contributes to its own regulation throughout embryonic patterning. For example, the gap gene *knirps* is 3 kb long while *knirps-related* is 23 kb due to numerous introns. *In vitro* experiments showed that Knirps-related can bind Knirps repressor sites. However only expression of the shorter Knirps-related cDNA, and not its endogenous gene locus, can rescue *knirps* mutant embryos (Rothe et al., 1992).

Similar experiments with the 21 kb gene *short-gastrulation* (*sog*) detected *in vivo* expression with probes only against the 5' region and not the 3' end of the transcript (Reeves et al., 2012). Perhaps increasing transcript length through introns is a mechanism to prevent premature function. Conversely, genes with shorter transcripts are ensured to have a functional protein during nuclear cycles as in the case for Snail. *snail* is less than 2 kb and is a transcription factor with inputs into multiple early genes. Its repressive activity is observed in nc 13 and must be re-established at the beginning of nc 14 (Reeves et al., 2012). The cooperation between gene length and short nuclear cycles appears to be a general method to regulate gene expression. However, patterning also requires spatial control in addition to the temporal dynamics of nuclear cycles.

*Spatial considerations in gene patterning*

In nuclear cycles 1-13, the nuclear envelope breaks down during each division and nuclei reach a local equilibrium with the surrounding cytoplasm (DeLotto et al., 2007). This hinders the extent to which transcripts and other factors can be spatially regulated. Cellularization into roughly 6,000 cells during nc 14 allows a tighter control in the position of expression domains.

The maternal to zygotic transition (MZT) begins in nc 13 and continues throughout nc 14. During this period, there is a sharp degradation of maternal product concurrent with a burst of zygotic transcription. MZT is distinct from the mid-blastula transition (MBT) when a gap phase, G2, is introduced in nc 14 after synthesis but before the $14^{th}$ division (Lee and Orr-Weaver, 2003). By this point, nc 14 persists for 50 minutes, allowing the embryo time to cellularize. Cytoskeletal elements form furrow canals and eventually the

basement membranes connect neighboring cells (Mazumdar and Mazumdar, 2002). These series of events often occur in parallel and are sometimes not distinguished as separate processes (Tadros and Lipshitz, 2009) (Figure 1).

There have been many studies on the exact inputs into these transitions. Previous models such as a fixed time after fertilization or counting cell cycles have been negated (Lu et al., 2009). One prevailing view is the change in nucleocytoplasmic (N/C) ratio, or the concentration of DNA, that prompts MBT/MZT and also perhaps the slowing down of nuclear cycles 10-13 (Edgar et al., 1986; Lu et al., 2010; Sibon et al., 1997). There are, however, exceptions of genes that do not depend on the N/C ratio and in fact support MBT/MZT (Lu et al., 2009; Sokac and Wieschaus, 2008; Sung et al., 2013). Maternal genes have a major role in supporting gene patterning and initiating MBT/MZT. With the consecutive nuclear divisions of the syncytium, there is a temporal limit to the extent that transcription factors can activate gene expression. It is therefore maternal mRNAs that contribute to the robustness of gene patterning in the early embryo. For example, many of the key transcription factors required for establishing the dorsal-ventral and anterior-posterior axes are deposited maternally (discussed below). These maternal genes are also important to initiate zygotic transcription when the short nuclear cycle intervals no longer inhibit completion of transcripts. However, expression of maternal products is often ubiquitous, which limits their spatial influence. Hence, the formation of cellular membranes often occurs in conjunction with zygotic transcription, which can be spatially regulated.

Gene patterning initially starts with a few factors, usually maternal, which establish a set of secondary regulators that in turn continue to initiate differential expression (Figure 1, green arrows). These inputs are combinatorial and the additional layers of information

allow precise spatial domains (Spitz and Furlong, 2012). For example, the transcription factor Zelda is maternal and zygotic, and both contributions are required to support its full activity as a global activator. In addition to regulating cellularization genes, Zelda is involved in sex determination and axis formation, indicating its broad function (Liang et al., 2008; Liberman and Stathopoulos, 2009). Because Zelda is ubiquitous, it alone cannot provide spatial information for patterning. Therefore, it follows that additional factors are employed to define domains of gene expression. Analysis of enhancer regions, DNA sequences that support transcription, reveal binding sites for multiple factors including Zelda (Spitz and Furlong, 2012). Below, we focus on the spatiotemporal dynamics and combinatorial inputs required to establish the dorsal-ventral (DV) axis of the *Drosophila* embryo.

*Formation of the dorsal-ventral axis*

The Dorsal morphogen is a transcription factor pivotal in patterning the DV axis. Classical models of morphogen gradients describe a graded distribution across a field of cells that, at specific threshold levels, will spatially divide the embryo into different domains of differentiation. Dorsal is maternally provided and forms a gradient such that its highest concentration is in the ventral-most nucleus while being excluded from dorsal nuclei. It regulates gene expression, traditionally classified into Type I-III Dorsal target genes, in domains along the DV axis (Reeves and Stathopoulos, 2009). The boundaries of these domains range from sharp to graded, resulting in many investigations on how Dorsal is able to produce such patterns. In addition, there have been questions regarding Dorsal's

range of action along the DV axis where its concentration is low, and yet cells in these lateral regions of the embryo are still able to read threshold levels.

Advances in imaging technology have allowed us to probe into the spatiotemporal dynamics of Dorsal and how its gradient is established. Using two-photon light-sheet microscopy (2P-SPIM), we found that Dorsal levels never reaches steady state. Over nuclear cycles 11-14, Dorsal concentrations increase in ventral nuclei while decreasing in dorsal nuclei (see Chapter II for details). Another study reports active shuttling of the Dorsal protein in and out of the nuclei (DeLotto et al., 2007). Several models are possible to describe this delivery, and further investigation is required to determine the mechanism Dorsal is allocated (see Chapter VII Discussion). Nevertheless, not all morphogen spatial domains are generated in the same fashion as the Dorsal gradient.

Morphogen gradients can be established by a variety of mechanisms to allow for flexibility in its own regulation and its regulation of downstream genes. Another well-studied system is the Bicoid (Bcd) morphogen, which patterns the anterior-posterior (AP) axis in *Drosophila*. In contrast to Drosal, Biocid mRNA is differentially restricted to the anterior of the embryo where it is locally translated (Little et al., 2011). Moreover, the dynamics of Bicoid are unlike Dorsal: Bicoid was found to reach peak levels as early as nc 10, which did not significantly change throughout subsequent divisions suggesting some form of gradient stability (Figure 2) (Gregor et al., 2007b). However, one study shows that the Bicoid gradient is nonessential during nuclear cycles 11-13 (Lucchetta et al., 2008). This may be a consequence of cytoplasmic movements during the syncytial divisions that delays readout until nc 14 when the cytoplasm becomes partitioned into cells. Another possibility is that the Bicoid gradient is interpreted earlier in development through a form

of memory (additional examples discussed below). The stability of Bicoid may indicate its lack of requirement in these earlier nuclear cycles since there is no new information to convey. During nuclear cycles 11-13, Bicoid-dependent patterning could have already initiated cross-repressive interactions between transcription factors. Therefore, AP patterning is not governed by Bicoid itself at this time, but by its target genes. While it is yet to be determined if the Dorsal gradient is required prior to nc 14, initial analysis implies that its spatiotemporal dynamics contribute to expression of its downstream genes (Reeves et al., 2012; Trisnadi et al., 2013). Other morphogen gradients give rise to patterns through cellular memory.

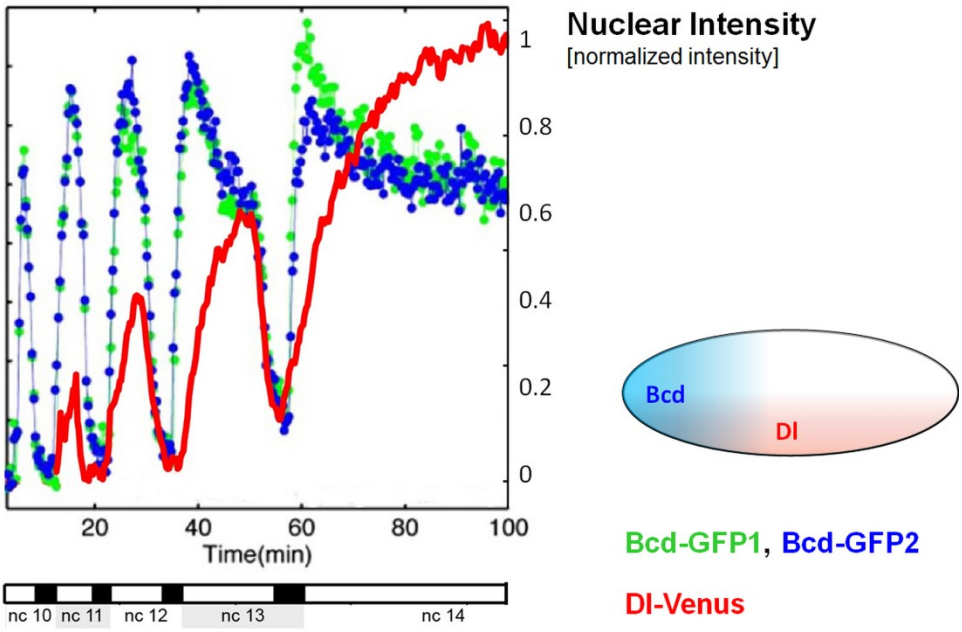



**Figure 2. Comparison of the Dorsal and Bicoid morphogens.** Bicoid (green, blue) concentration reaches peak levels by early nuclear cycles (nc) and does not change. In contrast, Dorsal (red) levels build with each successive nuclear cycle. Schematic embryo showing the two morphogen gradients; anterior is left and ventral is down. Bicoid data from Gregor et al., 2007b; Little et al., 2011; Dorsal from Reeves et al., 2012.

Evidence suggests that the duration of morphogen signaling can be necessary for proper patterning as well as its concentration (Alexandre et al., 2014; Dessaud et al., 2010). This has been mathematically predicted through modeling of the patterning network in the *Drosophila* wing disc with the Hedgehog ligand and Patched (Ptc) receptor. The network is structured such that Hedgehog signaling increases *ptc* expression by inhibiting Ptc repression. Experiments show that the Ptc domain initially overshoots and then later refines its boundary. When the network architecture was altered, Ptc could no longer recede its expression domain and remained expanded (Nahmad and Stathopoulos, 2009). This memory-like mechanism can act after the gradient has withdrawn and downstream gene domains refine in accordance to a cell's history of exposure to the morphogen. As the embryo continues to develop, morphogens may be modified – degraded or transported, for example – before its downstream activity is measured. As previously discussed, short timescales may require a form of regulation that account for signal propagation even when the immediate output is postponed. In light of the observed dynamics in the Dorsal gradient, we analyzed the expression domains of its target genes over multiple nuclear cycles and used modeling of the Dorsal network to calculate patterning responses.

Further examination of carefully staged fixed embryos revealed that genes downstream of Dorsal shift their boundaries, particularly during nuclear cycles 13 and 14. These borders are seen to refine from graded to sharp and domains expand from narrow to broad. Computational simulations correlate these changes with the Dorsal dynamics observed (see Chapters II and III for details). While gene length can regulate the timing of activity, as previously discussed, this property can also be indirect. For example, transcriptional elongation continues even when Snail-mediated repression inhibits *de novo*

initiation (Bothma et al., 2011). This results in persistent mRNA production of larger transcripts such as *sog*, since all initiated polymerases will complete transcription. Smaller transcripts will have fewer initiation events after repression due to its size. This delay in repression predicts graded, not sharp, borders due to the transiency of these longer genes (McHale et al., 2011). Snail was also found to expand its domain as Dorsal levels increase since, presumably, more cells reach threshold levels to activate its expression (Reeves et al., 2012). The short length of Snail allows refinement within nc 14, highlighting the changes on minute timescales. In these instances, information regarding transcript size may be integrated into the network design through number of introns and regulatory binding sites in *cis*-regulatory modules (CRMs). The architecture of this network is also important to define the spatiotemporal dynamics of downstream gene responses, as shown for Hedgehog patterning. CRMs often receive multiple inputs from a variety of transcription factors that dictate gene expression. Modeling suggests that the additive layers of control reduce the stochasticity and increase the spatiotemporal specificity of expression domains to result in robust patterning (Gregor et al., 2007a; Reeves et al., 2012).

There have been multiple studies showing combinatorial inputs into genes patterning the DV axis. As previously mentioned, Zelda is a known global activator and functions with the spatially restricted Dorsal to support specific domains of gene expression (Liberman and Stathopoulos, 2009). In addition, several downstream Dorsal target genes, such as Twist, provide spatial information and are required to work with Dorsal to activate various gene expressions (Reeves and Stathopoulos, 2009). Repressors also act to refine domains, such as Snail on *sog* (discussed above). Another example of tiered regulation is the dorsal-lateral gene *intermediate neuroblast defective* (*ind*), which is thought to have

negative inputs from TGF-β signaling and the Capicua repressor as well as to receive positive inputs by Dorsal and EGFR signaling (Garcia and Stathopoulos, 2011). These levels of regulation allow positional correction over time that results in precise patterning when a single factor may be individually less spatially accurate. Current investigations are aimed at isolating the various contributions into genes of interest.

Many genome-wide chromatin immunoprecipitation (ChIP) studies have allowed us to identify binding of transcription factors to DNA sequences at various time points (Ozdemir et al., 2011; Sandmann et al., 2007; Zeitlinger et al., 2007). This does not necessarily mean that each binding event detected is functional (i.e., that it produces a transcriptional output). However, even though such enhancer sites should be confirmed, this has allowed us to realize the multiple regions required for a transcriptional event and the complexity of enhancer mechanisms (Spitz and Furlong, 2012). It has previously been shown that for many loci, such as the segmentation gene *eve*, there are multiple enhancers that individually produce discrete expression domains. When combined, these independent enhancers produce the full pattern associated with the gene. This design of multiple, seemingly autonomous modules also appears to hold true for dorsal-ventral genes, though it was not previously appreciated (Dunipace et al., 2011; Kvon et al., 2014). These expression domains may abut each other and are not as apparent as with the case of segmentation genes. Enhancers may also appear redundant, but their collective requirement becomes apparent at different stages or during developmental stress such as temperature. Therefore, careful spatiotemporal analysis to link enhancers with specific patterns is required.

Post-transcriptional regulation is another mechanism of gene patterning. Similar to the regulation of maternal products, mRNA localization has been found to be

spatiotemporally dynamic in some cases (Lecuyer et al., 2007). Early onset of transcription can be visualized by their punctate nuclear localization, and then later transcripts are usually diffuse throughout the nucleus and cytoplasm. However, for some genes, their mRNA is distributed on the apical side at the end of a nuclear cycle (Reeves et al., 2012). It is unknown if any translation occurs at these sites. mRNAs may also encounter stabilizing or destabilizing factors that control their post-transcriptional expression. Another mechanism used during development is to remove cells with improper expression. A targeted cell death mechanism is used to repair the Bicoid gradient in late stage embryos (Tanaka et al., 2014). As more forms of regulation are characterized, it is likely that the spatiotemporal dynamics of gene expression will continue to be significant in patterning.

**Spatiotemporal regulation of signaling pathways**

The onset of gastrulation marks the separation of germ layers in the embryo, which has already been differentially patterned into a cell fate map. Cells now start to adopt characteristic physical shapes in addition to producing different signals. These feed back into gene patterning as development continues.

Combinatorial inputs include signaling pathways that are also spatiotemporally regulated, and this information is passed on to their target genes. For example, expression of *single-minded* (*sim)*, which derives the ventral midline of the embryo, is activated by initially broad Notch signaling. However, Notch itself is then carved out by Snail repression to form the ventral boundary and inhibitors from the dorsal regions. This results

in a narrow *sim* domain by the end of nc 14 (Cowden and Levine, 2002). Another case is the expression of *zerknullt* (*zen*) in the amnioserosa through Dpp signaling during cellularization. Dorsal and Zelda together activate *zen* in a broad dorsal domain during nuclear cycles 11-13 (Liang et al., 2008). However in nc 14, *zen* expression becomes Dpp-dependent. At the same time, Brinker (Brk) repression extends from the ventral margins and competes with Dpp for binding sites in *zen* enhancers (Rushlow et al., 2001). This theme of combinatorial inputs to regulate gene expression continues throughout development. A blend of the various types of contributions takes effect during mesoderm migration.

*Early development of the mesoderm*

After cellularization at nc 14 is complete, the presumptive mesoderm undergoes epithelial-to-mesenchymal transition (EMT). One hallmark of EMT is the downregulation of adhesion molecules E-cadherin (Ecad) and upregulation of N-cadherin (Ncad) (Baum et al., 2008). Dorsal targets Snail and Twist contribute to the repression of Ecad (Oda et al., 1998) and the activation of Fibroblast growth factor (FGF) receptor Heartless (Htl) (Shishido et al., 1993) in the mesoderm. Preliminary analysis of the network between Ecad, Snail, and FGF signaling suggests that FGFs may modulate Ecad through inhibition of *snail* expression. Subsequent migration of the mesoderm involves specific adhesive properties to balance between motility and association with the collective. Multiple inputs to regulate adhesion would provide tight spatiotemporal control of cadherins (see Chapter VI for details).

FGF signaling may also direct mesoderm migration independently of Ecad, though other yet to be identified pathways are thought to be involved as well. Mesoderm migration can be described as a multistep process that includes the EMT and invagination, collapse of the ventral furrow, dorsal spreading, and intercalation to form a monolayer (Figure 3). FGF signaling was found to direct furrow collapse and intercalation, both in the radial direction (McMahon et al., 2010). In contrast to Htl expression in the mesoderm, the two FGFs (Pyramus and Thisbe) are expressed in the ectodermal substratum (Stathopoulos et al., 2004). This arrangement of complementary patterns results in the ligand-expressing ectoderm guiding the receptor-expressing mesoderm in a radial movement (see Chapter IV for review).

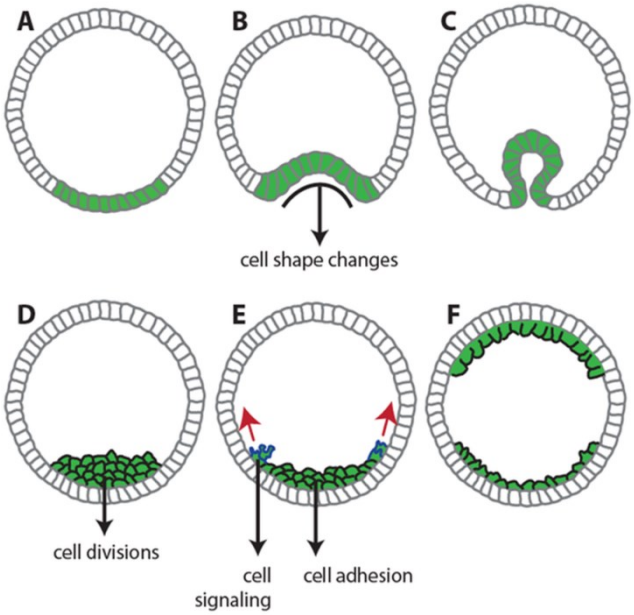



**Figure 3. Migration of the mesoderm collective.** Representative cross-section of *Drosophila* embryos staged 5-10. (A) The presumptive mesoderm is defined. (B) Invagination of the mesoderm occurs during EMT. (C) The mesoderm forms a ventral furrow. (D) The tube collapses upon the ectoderm in an FGF-dependent process. (E) Mesoderm cells spread laterally. (F) Intercalation requires FGF signaling and the mesoderm forms a monolayer. Green indicates mesoderm cells which express the FGF receptor; gray indicates the ectoderm with FGF ligands; ventral is down.

Despite the successful efforts to express signaling factors at the correct time and place, there are often additional post-translational forms of regulation. Many pathways work in conjunction with co-factors to influence their function in a spatiotemporal manner. FGF-FGFR complexes interact with heparan sulfate proteoglycans (HSPGs) to support signaling propagation (Lin, 2004). Comparison of two HSPGs, the secreted Terribly reduced optic lobes (Trol) and the membrane-bound Syndecan (Sdc), revealed expressions that are spatiotemporally dynamic during mesoderm migration. Trol, which is an extracellular-matrix (ECM) component, is first refined to express in ventral regions of the ectoderm prior to furrow collapse. It appears to facilitate long-range FGF signaling during migration and mediate radial movements at a distance. On the other hand, Sdc is restricted to the membrane and was found also in ventral regions of the ectoderm but in later stages after lateral spreading is complete. It is required for signaling between neighboring cells as in intercalation, the final step of migration, and subsequent differentiation in mesoderm development. Their patterns depend on function such that they are differentially expressed to support FGF signaling in cells transitioning from migration to differentiation (see Chapter V for details).

**Setting the stage: considerations during morphogenesis**

Understanding the spatiotemporal dynamics of gene patterning becomes more challenging as new inputs are introduced into the system. In the process of forming germ layers, cells come into contact with multiple cells along the way. Depending on the context, cells may need to be able to segregate between same vs. different cell types or rely on other cells for

guidance. For example, directional cues by definition are only present in specific patterns in order to provide positional information, like FGF signaling. Other co-factors will accordingly have restricted domains of expression, as in the case for HSPGs. Protrusions add another layer of complexity as recent studies have shown that membranes containing signaling receptors can extend several cells away to receive signal from non-neighboring cells (Roy et al., 2014). As cells migrate, this spatial information may need to change at the proper time correspondingly.

The surrounding environment is also important for development. The ECM contains many factors that input into the cell. They range from survival elements to stabilizing ligands for signal diffusion (Kim et al., 2011). Other examples include more passive roles such as traction for cell mobility or barriers to prevent cell mixing. A migrating population of cells will encounter multiple landscapes that regulate their function in a spatiotemporal manner.

Mitosis no longer occurs synchronously after nc 14 and instead is driven by the domains delineated from differential gene expression (Foe, 1989). Once a cell divides, it retains information on its fate and other associated characteristics. Evidence suggests that sister cells in the mesoderm physically attach to each other even after brief periods of being separated by distance during migration (McMahon et al., 2008). Specific cell-cell interactions can be important for the spatiotemporal organization of a collective and for communication, as in cases of differentiation.

**Gastrulation through spatiotemporally restricted gene expressions**

Our interests in understanding how the embryo is patterned must take into account the changing environment inherent in the system and the combinatorial modes of regulation. Careful staging and measurements of gene boundaries are prevalent in our investigations. The following research quantified the spatiotemporal dynamics in the Dorsal morphogen gradient and correlated these changes with gene domain refinement through modeling. We also examined factors involved in mesoderm migration and found a role for proteoglycans in supporting FGF signaling during this process. In addition, this study extends the network between FGFs, Snail, and cadherins in the early mesoderm. Our survey into embryonic patterning highlights the importance of the role that gene expression dynamics has in gastrulation.

CHAPTER II

# Dorsal-ventral gene expression in the *Drosophila* embryo reflects the dynamics and precision of the Dorsal nuclear gradient

This chapter was written with Gregory T. Reeves, Thai V. Truong, Marcos Nahmad, Sophie Katz, and Angelike Stathopoulos, and published in *Developmental Cell,* 22(3)544-557 in 2012.

ABSTRACT

Patterning of the dorsal-ventral axis in the early *Drosophila* embryo depends on the nuclear distribution of the Dorsal transcription factor. Using live two-photon light-sheet microscopy, we quantified the nuclear Dorsal gradient in space and time and found that its amplitude and basal levels display oscillations throughout early embryonic development. These dynamics raise questions regarding how cells can reproducibly establish patterns of gene expression from a rapidly-varying signal. We therefore quantified domains of Dorsal target genes, discovering their expression patterns are also dynamic. Computational modeling of this system reveals a correlation between Dorsal gradient dynamics and changes in target gene expression and suggests that these dynamics, together with time-averaging of noise, results in the formation of graded gene expression borders in regions where the gradient is nearly flat. We propose that mRNA levels remain plastic during transient signaling events, allowing tissues to refine patterns in the face of genetic or environmental variation.

# INTRODUCTION

In a developing organism, tissues have long been proposed to be patterned by spatially graded signals that specify cell fate in a concentration-dependent manner. Classically, these "morphogens" have been defined as originating from a defined source and forming a graded distribution by diffusion and degradation; however, in recent years it has become clear that morphogens can become spatially organized by a variety of mechanisms. Two of the best-characterized morphogen gradients pattern the anterior-posterior (AP) and dorsal-ventral (DV) axes of the *Drosophila* early embryo: the Bicoid and Dorsal transcription factor gradients (rev. in Porcher and Dostatni, 2010; Reeves and Stathopoulos, 2009). Their graded distributions are established using very different mechanisms. Bicoid is locally translated because its mRNA contains a localization sequence; whereas, Dorsal is localized to the nucleus more strongly in the ventral regions of the embryo because of localized Toll-receptor associated signaling. Live imaging has revealed significant dynamics in the exact levels of Bicoid (e.g., Gregor et al., 2007b; Little et al., 2011); however, the dynamics of target gene expression examined in fixed embryos suggest that the levels of Bicoid are important, but not the only defining factor in the expression of target genes (e.g., Jaeger et al., 2004; Ochoa-Espinosa et al., 2009). In contrast, no study to date has investigated systematically temporal features of the Dorsal gradient and its relationship to the expression of its target genes.

The role of Dorsal in the expression of its target genes has been conceptualized as the concentration dependent activation of genes, divided into three broad categories (Types I, II, and III) based on both their domains of expression and their presumed threshold-

dependent responsiveness to Dorsal levels (see Figures 1A and 1B; rev. in Reeves and Stathopoulos, 2009; Stathopoulos and Levine, 2005). Type I genes, such as *twist* (*twi*) and *snail* (*sna*), are expressed in ventral regions of the embryo in a domain where the levels of nuclear Dorsal are high (up to ~20% DV position; where 0% is the ventral-most position and 100% is the dorsal-most position). Type II genes like *ventral nervous system defective* (*vnd*), are thought to be expressed in ventrolateral domains (dorsal boundaries at roughly 33% DV position) through the combined actions of enhancers that are of intermediate affinity to Dorsal and that are repressed by the Snail transcription factor in the ventral-most regions. Type III genes are expressed in domains with boundaries past 45% DV position, and can be further subdivided into two categories: those that are activated by Dorsal [Type III+, such as *short-gastrulation* (*sog*)] and those that are repressed by Dorsal [Type III–, such as *zerknüllt* (*zen*) and *decapentaplegic* (*dpp*)]. Presumably the lowest levels of Dorsal are sufficient to determine the spatial extent of Type III target genes, but the roles played by other factors remain unclear (e.g., Jiang and Levine, 1993; Liberman and Stathopoulos, 2009).

To study the role of nuclear Dorsal in controlling gene expression, a number of studies have attempted to measure nuclear Dorsal, using either antibody stainings in fixed tissues (Chung et al., 2011; Liberman et al., 2009; Zinzen et al., 2006) or imaging of Dorsal-GFP in live embryos (DeLotto et al., 2007; Kanodia et al., 2009). Our previous study in fixed embryos showed that the Dorsal gradient was more narrow than often described, resulting in a relatively flat distribution more than 110 microns from the ventral midline (40% DV position); this raised the question of how Dorsal could specify gene expression in this domain (Liberman et al., 2009). In contrast, others reported broader

gradients in live embryos, arguing that the Dorsal concentration could carry positional information up to at least 60% DV position; furthermore, they argued that the Dorsal nuclear concentration continued to decline all the way to the dorsal midline (compare dotted vs. solid black curves in Figure 1B; Chung et al., 2011; Kanodia et al., 2009). Both sets of studies suggested that the Dorsal nuclear gradient is dynamic, varying in time both within nuclear cycles and from one nuclear cycle to the next (DeLotto et al., 2007; Kanodia et al., 2009; Liberman et al., 2009).

Previous live studies of Dorsal nuclear concentration and dynamics (DeLotto et al., 2007; Kanodia et al., 2009) employed a Dorsal-GFP fusion that results in a measurably wider gradient than wildtype and also fails to fully complement *dorsal* null mutants (Liberman et al., 2009). Furthermore, these studies employed conventional confocal microscopy, in which nuclear motion and limited light penetration both complicated an accurate measurement of the Dorsal-GFP nuclear gradient in both time and space (DeLotto et al., 2007; Kanodia et al., 2009). Thus, the dynamics of the Dorsal gradient have not yet been satisfactorily measured, nor has it been investigated how these dynamics might impact domains of gene expression.

In this study, we employed a *dorsal-venus* fusion transgene and improved microscopy to address two outstanding questions regarding the Dorsal gradient: first, how does a highly dynamic morphogen signal specify gene expression domains, and second, how does a narrow gradient deliver precise positional information to the entire DV axis? We find that the Dorsal nuclear gradient varies in both time and space during nuclear cycles (nc) 11-14, and that the expression of Dorsal target genes is often as dynamic as the gradient. Furthermore, we suggest that the graded boundaries in the expression patterns of

Type III genes result from a time-integration of the nearly-flat gradient tail. We used a threshold-based model to show that much of the dynamics and sharpness of Dorsal target gene expression patterns can be accounted for by the dynamics and shape of the Dorsal nuclear gradient.

# RESULTS

## *Use of a Dorsal-Venus fusion to monitor Dorsal in living embryos using light-sheet microscopy*

To create a transgene encoding a fully functional Dorsal-Venus fusion protein, we BAC-recombineered 25 kb of genomic DNA surrounding the *dorsal* locus with sequences encoding the *venus* yellow fluorescent protein optimized for *Drosophila,* inserted in-frame at the C-terminus of the Dorsal protein (Materials and Methods). This *dorsal-venus* transgene fully complements null mutations in *dorsal* when present at one copy, similar to an unmodified *dorsal* rescue transgene. In contrast, neither the 25 kB *dorsal-gfp* construct we constructed nor previous *dorsal-gfp* cDNA based constructs complement the *dorsal* mutant at one copy (Liberman et al., 2009; see Materials and Methods). Live imaging and immunostaining demonstrate that Dorsal-Venus exhibits a distribution more similar to wildtype Dorsal than Dorsal-GFP. We defined a quantitative measure of the width of the gradient by fitting the data to Equation 1 (see Materials and Methods), resulting in a metric of $\sigma$ (Liberman et al., 2009). This analysis shows that embryos carrying Dorsal-Venus 25 kB rescue construct have a width of $\sigma = 0.16 +/- 0.01$ (standard deviation), which is much more similar to the width of the wildtype Dorsal gradient ($\sigma = 0.14 +/- 0.01$) than that from embryos carrying a 25 kB Dorsal-GFP construct that we made ($\sigma = 0.20 +/- 0.02$) (see Figures 1C-F). The similarity of the Dorsal-Venus distribution to that of Dorsal offers us the opportunity to accurately assess the spatiotemporal behavior of functional Dorsal nuclear gradients using live imaging.

To quantitatively measure the levels of Dorsal-Venus in early embryos, we imaged embryos from mothers containing one copy of a *dorsal-venus* transgene, one copy of the endogenous *dorsal* gene, and one copy of an *H2A-RFP* transgene to label all of the nuclei so they can be unambiguously segmented (Materials and Methods). Using two-photon scanned light-sheet microscopy (2P-SPIM) (Figures 2A and 2B; Truong et al., 2011), which provides superior resolution at high sample depth compared to conventional confocal microscopy, we imaged end-on cross sections of the nc 14 embryo to determine the Dorsal-Venus nuclear gradient between 50 and 250 microns from the anterior pole. We found the gradient changes with AP location, becoming increasingly wider 100 microns or closer to the pole (Figures 2C-F). Therefore, our quantitative analysis of the Dorsal gradient in the rest of this study will be based on measurements made from images of H2A-RFP and Dorsal-Venus during nc 11-14 in optical cross sections of embryos 150 microns from the anterior pole (i.e., just posterior to the presumptive cephalic furrow; see Figure 3A and Movie S1).

*Dynamic properties of the Dorsal-Venus nuclear gradient*

We analyzed overall spatial properties of the Dorsal nuclear gradient throughout the course of nc 11-14 by collecting 2P-SPIM time lapse images. The image processing tools extracted the nuclear Dorsal-Venus signal by segmenting the nuclear regions based on the H2A-RFP images. The parameters of the gradient amplitude ($A$), basal levels ($B$), and width ($\sigma$) were determined using the Gaussian-fitting described in Materials and Methods, and each can vary over time (see Figure 3B and Equation 1 in the Materials and Methods).

The gradient amplitude [$A(t)$] increases from nuclear cycle to nuclear cycle (blue curve in Figure 3C), agreeing with previous data using fixed samples (Liberman et al., 2009) and predictions from modeling studies (Kanodia et al., 2009). Moreover, the Dorsal-Venus gradient amplitude exhibits a "saw-tooth" pattern over time, never reaching steady state, consistent with single-nucleus traces from other live studies (DeLotto et al., 2007). This pattern can be explained by the nuclei filling with Dorsal-Venus relatively slowly throughout each nuclear cycle interphase, then rapidly equilibrating with the cytoplasm when the nuclear envelopes break down at the beginning of mitosis.

In contrast to the filling of the ventral-most nuclei with Dorsal-Venus, the dorsal-most nuclei appear to begin each interphase with "too much" Dorsal-Venus. As interphase proceeds, these dorsal-most nuclei slowly evacuate nuclear Dorsal, causing the Dorsal levels in these nuclei [i.e., the "basal levels" of the gradient, $B(t)$; see orange trace in Figure 3C] to decrease during interphase. When mitosis begins, basal levels rapidly increase. This counter-action between the Dorsal levels building in the ventral-most nuclei and declining in the dorsal-most, fits with the notion that the nuclei begin each interphase with Dorsal levels equilibrated with the cytoplasm, and only after an intact nuclear envelope forms can selective nuclear import/export processes develop the nuclear concentration gradient. With the successive import of Dorsal in ventral nuclei during each syncytial cycle, a cytoplasmic Dorsal gradient also develops that can be seen by the end of nc 13 mitosis when the nuclear envelope breaks down (Figures 3D-I). It was previously unappreciated that the overall levels of Dorsal protein are non-uniform along the DV axis.

Our analysis of the Dorsal-Venus gradient over time shows a remarkably constant gradient width [$\sigma(t)$] across all time (interphase only, red trace in Figure 3C), implying the

Dorsal nuclear gradient always becomes nearly flat at the same location (roughly 40% DV position). This result was also suggested by our previous fixed tissue analysis, in which the gradient width appeared to be unchanging from nuclear cycle to nuclear cycle (Liberman et al., 2009). However, the previous results obtained with fixed tissues left open the possibility that gradient widths could change either subtly across nuclear cycles or significantly within a nuclear cycle. The live imaging done here dismisses that possibility.

Measurements of the Dorsal-Venus nuclear gradient in three live embryos revealed similar results (see Figure S1). Together, these results underscore two questions regarding the action of the Dorsal gradient. First, how does a constantly-changing morphogen gradient specify domains of gene expression? Gene expression patterns might be established early then depend on cis-regulatory action, as has been proposed for the Bicoid network (e.g., Bergmann et al., 2007; Jaeger et al., 2004). Alternatively, mRNA transcription of the target genes might constantly change, driven by the dynamic changes in nuclear Dorsal. Second, how does a signal as narrow as the Dorsal nuclear gradient control the expression pattern of genes past 40% DV position (e.g., the Type III genes)? In other words, how could Dorsal provide reliable positional information given that its gradient is nearly flat?

*Dorsal target gene expression patterns in space and time*

To address the first question, we examined gene expression patterns in manually cross-sectioned, wildtype embryos using multiplex in situ hybridization during nc 11-14 (see Figure 4 and Supplemental Experimental Procedures). As nuclear cycles get progressively longer, we divided nc 13 and 14 into early/late and early/mid/late timepoints,

respectively, based on nuclear morphology and density (Figure 4A). The brightness and contrast of images in Figure 4A-E were intentionally adjusted to visually highlight distinctions; whereas, the profiles in Figure 4F come from analysis of the raw images (see Supplemental Experimental Procedures). In order to obtain a representative sampling of target gene dynamics, we examined four classical Dorsal-target genes of Type I (*sna*), II (*vnd*), III+ (*sog*), and III- (*zen*) patterns. Genetic and cis-regulatory analyses support the view that these genes are Dorsal targets (rev. in Reeves and Stathopoulos, 2009).

Each of these classical Dorsal target genes have dynamic patterns of expression. *zen* is expressed in the dorsal half of the embryo as early as nc 11, and builds in time until mid-to-late nc 14, when its expression pattern refines into a narrow stripe (Figure 4E), presumably from Dpp signaling (Rushlow et al., 2001). On the other hand, the expression patterns of *sna*, *vnd*, and *sog* exhibit more complex dynamics (Figures 4B, 4C, and 4D, respectively). *sog* transcripts, nuclear-localized and likely nascent, are observed as early as nc 12. The mature (non-nuclear) mRNA for all three of these genes is first seen in nc 13 (Figures 4B-D; see also Figure S2A). During this nuclear cycle, both *sog* and *vnd* are initially expressed in ventral regions where *sna* normally would repress them. As nc 13 progresses, the levels of *sog* and *vnd* increase only outside the *sna* domain, presumably because increasing activity of Sna repression limits expression in ventral regions. At the onset of nc 14, *sog* and *vnd* patterns are present with more uniform expression in ventral and lateral domains; little evidence of Sna-mediated repression is apparent. However, as nc 14 continues, and the levels of both *sog* and *vnd* increase, ventral repression becomes apparent again.

Our data demonstrate that domains of gene expression change during the transition between nc 13 and nc 14, in that repression in ventral regions is no longer apparent at the start of nc 14 (Figure 4F and Figure S2). One possible explanation is that expression observed at the onset of nc 14 corresponds to early transcripts that avoid Sna repression, because Sna levels in early nc 14 are not high enough to repress *sog* and *vnd*.

To investigate further, we examined *sog* transcript localization more closely. At the beginning of nc 13, *sog* is present only as nuclear dots, likely sites of nascent transcription in the nucleus, while *sog* in late nc 13 is localized outside the nucleus near the apical membrane. We interpret non-nuclear *sog* transcripts as being complete mRNAs, as they appear to have been exported from the nucleus. At the onset of nc 14, *sog* is once again co-localized with nuclear staining and non-nuclear transcripts are absent (Figures 5A-C); non-nuclear transcripts appear again at mid nc 14 (Figure 4D). Moreover, the early nc 14 pattern shows little sign of Sna-mediated ventral repression. These results are consistent with the idea that the *sog* transcript is degraded in between nc 13 and nc 14.

We also inspected *sna* mRNA localization, thinking that if all transcripts are degraded at the nc 13/nc 14 transition, then perhaps this could explain the loss of Sna-mediated repression in ventral regions. Instead, *sna* transcripts exhibit a different trend: they switch back and forth between being delocalized (early nc 13, early nc 14) and apically localized (late nc 13, mid/late nc 14) (Figures 5D-F and Figure 4B). Little is known regarding the relationship between *sna* transcript localization/stability, but it has been observed that for a number of other genes, mRNA localization is thought to affect function (Lecuyer et al., 2007). We did not find evidence during our time-course that *sna* transcripts are completely degraded, as our results suggest for *sog*, nevertheless we

continued to test our working hypothesis, which was that decreased Sna levels result in derepression of genes at the start of nc 14.

We hypothesized that insufficient levels of Sna protein (rather than transcript) at the start of nc 14 might account for lack of ventral repression, therefore we examined levels of Sna protein within embryos relative to *sog* transcript. Embryos carrying a Sna-GFP rescue construct (Dunipace et al., 2011) were immunostained with anti-GFP, which is more robust than any anti-Sna antibody we have tried, and co-processed with a riboprobe to the 5' intron of *sog* (Figure S3A), which provides a near-real-time assay of active *sog* transcription. In analyzing these embryos, we found that Sna-GFP levels increase starting from nc 13 and reach peak intensity by the end of nc 14 (Figures 5G and 5I). While a slight dip in average Sna levels is observed between nc 13 and 14, the levels are highly variable, perhaps because the staging of our fixed embryos is not fine enough to capture the most rapid dynamics (e.g., Figure S2B). Nevertheless, in each embryo, we found a strong and consistent negative correlation between Sna-GFP and intronic *sog* in the ventral-most nuclei (Figure 5J and Figure S3B-F), suggesting that a threshold amount of Sna activity is required to extinguish *de novo sog* transcription.

*Lack of precision of the Dorsal gradient may explain graded expression profiles of Type III genes*

Our previous analysis of the Dorsal gradient led us to believe that its tails were flat, suggesting it could not provide the positional information necessary to specify the domains of expression of target genes such as *sog* and *zen* (Liberman et al., 2009). The live imaging of Dorsal-Venus performed here provided clear insights into Dorsal gradient dynamics, but

did not have enough signal to noise ratio to yield quantitative information on the gradient tail within the domain of low nuclear Dorsal levels. To circumvent this technical challenge and provide further insight into levels of nuclear Dorsal present in this domain, we quantified the Dorsal gradient in 153 fixed, manually cross-sectioned wildtype embryos at mid to late nc 14 (see Materials and Methods).

These measurements revealed that the Dorsal gradient displays two regimes (see Figure 6A). From the ventral midline up to 40% DV position, the gradient adopts a narrow Gaussian-like shape ($\sigma \sim 0.14$; see Figure 1F). More dorsally, however, the gradient becomes more linear and can be empirically characterized by a constant slope ($M$) multiplied by $x$ (the position along the DV axis; see Equation 2 in Materials and Methods). The average slope of the gradient tail (normalized by the gradient amplitude) is $-0.1 \pm 0.03$ (95% confidence interval; see histogram in Figure 6B). Thus, on average, nuclear Dorsal levels slowly decrease with increasing $x$. However, there is a considerable amount of variance in the distribution as 25% of embryos were measured to have a positive slope to the tail. While a portion of this variance reflects measurement error, error cannot account for the mean slope being definitively negative (see Figure S4).

With a gradually sloping tail and ventral, narrow Gaussian, it seems questionable that Dorsal could deliver precise positional information to lateral and dorso-lateral positions. To address this quantitatively, we evaluated the relative difference in Dorsal concentration ($\Delta c/c$) that would be seen by neighboring nuclei (see Equation 4 in Materials and Methods). In the gradient tail, at the dorsal border of *sog* ($x$ = 50% DV position), the difference in Dorsal levels that adjacent nuclei see is less than one percent (Figure 6C).

This is problematic as previous work has suggested that it is unlikely nuclei can reliably interpret concentration changes of less than 10% (e.g., Gregor et al., 2007a).

Another approach to examine the expected imprecision of the gene expression boundaries was to determine the effects of stochastic fluctuations in reading a shallow Dorsal gradient. In Figure 6D, we plot the Dorsal concentration for 40 nuclei along the semi-circumference of DV axis (using Equation 2 in Materials and Methods) and increase and decrease this amount by 10%, the level of read error suggested by previous studies (red curves in Figure 6D; Gregor et al., 2007a). With these fluctuations, the error in $x$ for placing a gene expression boundary outside the steep Gaussian-like regime is six or more nuclei, even with a gradient tail that reliably slopes downward. The implication is that Type III genes are located in a region where it is difficult for Dorsal to specify sharp, precise boundaries.

To investigate this issue further, we performed in situ hybridization of manually cross-sectioned nc 14 embryos with the antisense riboprobes of Type III transcripts. This analysis revealed that, in general, Type III genes (e.g., *sog*, *ths*, *Neu3*, and *zen*) possess graded borders at different DV positions (see Figures 6E and 6F). Perhaps the lack of precision in positional information results in the graded borders of these genes in dorsal regions (see Figures 6G and 6H). In other words, the noisy Dorsal gradient will activate gene expression in all nuclei in the graded border, but some more frequently than others depending on whether the nuclei read Dorsal to be above threshold. This would lead to a time-averaging mechanism in which mistakes are smoothed out as mRNA accumulates (see Figures 6G and 6H, and Figure S5) (Tostevin et al., 2007).

*Simulations of Dorsal-dependent patterning*

Live imaging of the Dorsal-Venus nuclear gradient has revealed complex dynamical behavior, with the gradient amplitude increasing and the basal levels decreasing over time. In addition, carefully examining mRNA expression in cross-sectioned embryos revealed graded boundaries of the Type III genes as well as gene expression dynamics across and within nuclear cycles. Together, these observations seem related, yet it is not intuitively obvious what quantitative effect the Dorsal gradient dynamics may have on the gene expression patterns. To test the plausibility of a causal relationship between the spatiotemporal dynamics of the gradient and its targets, we constructed a model of the wildtype Dorsal gradient based on our live and fixed tissue data (see Equation 2 and Materials and Methods). Moreover, we formulated a threshold-based model of mRNA dynamics to simulate expression patterns of *sna*, *vnd*, *sog*, and *zen* over nc 11-14, according to the network depicted in Figure 7A (see Equation 3 in Materials and Methods). In our model, the mRNA lifetimes and the thresholds that dictate gene transcription were fit such that the simulations would optimally match the experimental data shown in Figure 4F (see Figures 7C and 7D, and Materials and Methods).

The Dorsal levels simulated in space and time are shown in Figure 7B, in which the black curves are contours of constant Dorsal levels, corresponding to the fitted thresholds for gene expression (Type I, II, and III from left to right). These threshold contours suggest that gene expression boundaries will move in time as a result of the dynamics of the Dorsal gradient. This is demonstrated more clearly in Figure 7F, which depicts simulated Dorsal gradients near the end of nc 11-14 (horizontal arrows correspond to the signaling thresholds). The signaling threshold for Type II genes is located near 33% DV position

throughout all four nuclear cycles, even in the face of the dynamics of the Dorsal gradient (Figure 7F; see also Figures 7J-L), while the DV positions of Type I and Type III genes change significantly over nuclear cycles (red and green arrows in Figure 7F). In particular, during nc 14, the movement of the thresholds predicts Type I genes to expand, Type II genes to remain fixed, and Type III+ genes to begin the cycle expressed even in the dorsal-most nuclei (Figure 7E).

Simulations of gene expression patterns agree with these general predictions of threshold-dependent patterning. During nc 14, the simulated *sna* boundary moves dorsally (Figure 7G), the *vnd* boundary does not move (Figures 7J and 7K), and the *sog* domain begins broad, then the dorsal portion retracts (Figures 7M and 7N). These predictions prompted us to investigate the nc 14 dynamics of these three genes in more detail using fixed embryos. The *vnd* dorsal boundary remains static in early and mid nc 14, and expands only slightly in late nc 14 (Figure 7L). However, *sna* and *sog* expression is dynamic. As predicted based from our simulations, we found the *sna* domain expands during nc 14 (Figure 7I); and this specific result is supported by another study published recently (McHale et al., 2011). Furthermore, while *sog* mature transcripts are expressed in a constant domain (data not shown), importantly the boundary identified using an intronic *sog* probe, which serves as a 'real-time' proxy for responsiveness of transcription, shows the pattern retracts (Figure 7O).

Additionally, the threshold-based simulations offer a plausible explanation for the observed on/off cycling of Sna activity and its consequences on the ventral repression of *sog* and *vnd*. Simulated *sna* is expressed strongly in late nc 13, but decreases during nc 13 mitosis simply because transcription ceases while degradation continues (Figure 7H).

Afterwards, *sna* levels must build again in early nc 14 before it can repress *sog* and *vnd*. In the intervening time, *sog* (and *vnd*; not shown) is transcribed in the ventral-most nuclei (Figure 7H).

We identified several differences between the simulations and the observed patterns of gene expression, most notably that *sog* expression is difficult to accurately simulate. While the simulations correctly predict some aspects of the dorsal border of *sog*, in that it is graded and placed past 40% DV position (as explained by Figures 6G and 6H, and Figure S5), the simulated border is more graded than seen experimentally; the entire dorsal portion of the embryo expresses *sog* strongly. If the threshold parameter for *sog* were raised slightly higher to attempt to restrict *sog* expression more ventrally, only a very narrow final domain of *sog* is present (data not shown). The differences we see between the simulations and analysis of fluorescent in situ hybridization experiments may stem from a variety of reasons, including roles for other activators and repressors in supporting expression. For example, if a dorsally-acting factor (such as *zen* or another gene expressed in a similar domain) were to repress *sog*, then the lower threshold combined with the action of such a dorsally-acting repression could support expression of *sog* in a domain more comparable to the endogenous pattern (see Figure S5B).

Although Dorsal is an important player in patterning of these genes, there are indeed other inputs required for full DV patterning; for example, as stated above, some additional input is required to explain the *sog* dorsal boundary. Nevertheless this simple model incorporating only the interactions in Figure 7A does remarkably well, in that the model was able to demonstrate the plausibility that the observed gene expression dynamics is driven by the Dorsal gradient dynamics. In particular, the model successfully predicted

that Type I patterns expand during interphase, Type II patterns remain static, and Type III+ patterns begin nuclear cycles broadly expressed then retract.

# DISCUSSION

The observations that morphogen gradients are dynamic have raised questions about the influence of time on pattern formation (rev. in Kutejova et al., 2009). In this paper, we investigate quantitatively how the nuclear distribution of Dorsal, which rapidly changes throughout the nuclear divisions in the *Drosophila* blastoderm, gives rise to precise gene expression patterns. Our findings reveal that, in contrast to Bicoid, whose nuclear distribution stabilizes relatively quickly both between and within nuclear cycles (Little et al., 2011; Lucchetta et al., 2008), the Dorsal gradient is highly dynamic, exhibiting a temporally oscillating pattern of nuclear Dorsal concentrations that never reaches a steady state. The dynamics within a nuclear cycle result from the slow net nuclear import of Dorsal throughout each nuclear cycle interphase, followed by an abrupt export of Dorsal when the nuclear envelopes break down at the beginning of mitosis. We suggest these slow dynamics associated with nuclear localization of Dorsal relate to Toll-mediated signaling being required for its ability to gain competence to enter the nuclei. In addition, early nuclear cycles may concentrate an initially uniform distribution of Dorsal onto the ventral side of the embryo thereby redistributing the overall concentration of Dorsal protein over time.

The dynamics of the nuclear Dorsal gradient appear to determine the temporal evolution of gene expression. Our observations show that Dorsal target gene expression follows a dynamic pattern similar to the Dorsal gradient, both within and across nuclear cycles. An implication of these observations is that gene expression patterns are able to switch their on/off state in response to changes in the concentration of Dorsal. This is

similar to the manner that target genes respond to Activin in the Xenopus embryo (Gurdon et al., 1995). However, in contrast to Activin-dependent patterning, Dorsal target genes do not appear to exhibit a "ratchet effect", as it is possible to turn on/off gene expression by changing the levels of Dorsal above/below an activation threshold. Thus, the response of target gene expression to Dorsal levels appears to be a "real-time" response. Our data further suggest that the activity of Sna protein is also transient and/or that repression is delayed compared to onset of transcription, as both *sog* and *vnd* are ventrally-repressed by the end of nc 13 and derepressed in some early nc 14 embryos.

The highly dynamic patterning of genes along the dorsal-ventral axis documented in this study could possibly allow for fine-tuning of gene expression patterns to respond to feedback and/or buffer against genetic and environmental perturbation. In support of this hypothesis, many genes expressed at this stage along the DV axis support relatively short transcripts of less than 5 kB (such as *sna*, *twi*, *vnd*, *rho*, *brk*, *wntD*, and *zen* to name a few), and thus are able to respond quickly to changes in the Dorsal gradient and/or in cis-regulation of other DV genes. In contrast, genes with larger transcripts, such as *sog* or *Neu3* of 20+ kB in length, will take 15+ min to transcribe at a rate of 1.1-1.4 kB/min (Shermoen and O'Farrell, 1991; Thummel et al., 1990). Because any incomplete nascent transcripts are most likely aborted and degraded upon cell division (Rothe et al., 1992; Shermoen and O'Farrell, 1991), these long transcripts are particularly constrained by the rapid (~10 minutes) mitotic cycles of the early embryo. In addition, the action of Sna repression through transcriptional inhibition could be delayed in genes with long transcripts (McHale et al., 2011). Thus, transcription length can impact a gene's response to other factors (e.g.,

Sna) and, together with Dorsal, is proposed to play an important role in regulating the timing of gene expression.

The nuclear distribution of Dorsal does not reach a steady state and yet the expression patterns of most Dorsal target genes appear to stop changing at mid-to-late nc 14. This may result from the fact that the Dorsal gradient changes more slowly during this period and/or that at this point patterning may have been stabilized by the logic of the cis-regulatory network. At the onset of gastrulation, when Dorsal levels plummet, some patterns are extinguished whereas others are retained. We suggest those that are retained must rely on a Dorsal-independent mechanism to support expression along the DV axis. For example, several genes (e.g., *sim* and *vnd*) switch to autoregulatory feedback mechanisms to retain expression within the same domain and thereby no longer depend on Dorsal to support activation (Nambu et al., 1991; Von Ohlen et al., 2007).

The low slope of the Dorsal gradient in the lateral and dorso-lateral regions of the embryos makes it unlikely that the Dorsal morphogen gradient can specify precise domains of gene expression (i.e., sharp boundaries) here. Each Type III gene analyzed in this study exhibited a graded border, and our results lead us to propose two mechanisms that may contribute to this pattern. First, while a recent study showed that stochastic gene expression along the dorsal-ventral axis relates to polymerase pausing (Boettiger and Levine, 2009), we highlight that stochastic expression is a common phenomenon associated with most genes of Type III pattern. In addition, a time-averaging mechanism (Tostevin et al., 2007), we propose, will give rise to a graded expression response at the gradient tails. Second, if the basal levels of the Dorsal gradient decrease within a nuclear cycle, then the location where the Dorsal gradient crosses the putative Type III threshold will retreat from the

dorsal midline to roughly 50% DV position as demonstrated here with intronic *sog*. In this case, assuming transcripts are stable within a nuclear cycle, nuclei that transiently saw enough Dorsal to express the gene for only a given time window will be part of the graded domain. In all, our data suggest that both noise and dynamics may be factors contributing to proper patterning of genes beyond the spatial range of a morphogen. The graded nature of Type III patterns may influence their functions; for instance in the case of *sog* to support an inverse gradient of TGF-β signaling (e.g., Dorfman and Shilo, 2001).

As both the Dorsal gradient as well as its target genes change in time, this suggests a correlation between Dorsal levels and gene expression dynamics. Our model demonstrates that Dorsal gradient dynamics can plausibly account for the observed expression patterns in nc 13 and 14 for *sna*, *vnd*, and *zen*, capturing the general oscillatory nature of DV gene expression and, in particular, provides insight that ventral patterns expand and more dorsal patterns retract. However, the failure of the simulations to reproduce the dorsal border of *sog* could be explained by a missing component to the modeling caused by our limited understanding of the process, such as a dorsally-acting repressor. This would be consistent with other patterning systems in which cross-repressive interactions between target genes are important factors (Jaeger et al., 2004). Alternatively, the behavior of the dorsal border of *sog* could be explained by the additional input of activators such as Zelda (Liberman and Stathopoulos, 2009) or by a gradient tail that is steeper than our measurements suggest. Therefore, while genes exhibit dynamics in their expression that generally correlate with changing Dorsal levels, there is clearly more to understand.

It is becoming increasingly clear that steady state models of morphogen gradients ignore crucial developmental events and that modeling of systems that takes into consideration the dynamics is informative (e.g., Bergmann et al., 2007; Bolouri and Davidson, 2003; Jaeger and Reinitz, 2006; Kutejova et al., 2009; Lek et al., 2010). In some cases, the dynamics of morphogen gradients are instrumental in the establishment of "memory-like" patterns (Dessaud et al., 2010; Nahmad and Stathopoulos, 2009). In contrast, the on/off cycling of gene expression associated with the Dorsal system demonstrates plasticity rather than memory. We surmise this plasticity may be a critical design feature for the subtle fine-tuning of gene expression domains, or early initiation of genetic regulatory pathways that must operate in the short developmental time period of the *Drosophila* blastoderm.

# MATERIALS AND METHODS

*Construction of the dorsal-venus construct*

The 25 kb *dorsal-venus* and *dorsal-GFP* transgenes were generated using recombineering mediated gap repair performed using SW105 cells as previously described (Venken et al., 2006). The BAC encompassing the *dorsal* gene (BACR07M13) was obtained from the BacPac Resource Center and the attB-P[acman]-ApR was modified to contain approximately 600bp homology arms to the region of interest. Seamless insertion of *venus* or *gfp* just before the stop codon of *dl* was performed using the galK system (Warming et al., 2005). A 6XGly sequence was added before the start of both the Venus and GFP sequences using PCR. The final constructs were isolated and electroporated into EPI300 cells (Epicentre) and the copy number was induced using Fosmid Autoinduction Solution (Epicentre) according to the manufacturer's instructions. The constructs were isolated using Nucleobond EF plasmid midi prep kits (ClonTech) and injected into line 23648 (BDSC) at a concentration of 0.5-1µg/µl in water using standard techniques. All primers used for gap repair and recombineering are described in Supplemental Experimental Procedures.

These *dorsal* constructs fused to a fluorescent protein were inserted into the 86Fb landing site on the third chromosome (Bischof et al., 2007; Groth et al., 2004) and crossed into $dl^{1}$ and $dl^{4}$ mutant backgrounds (Bloomington stock center) to assay for ability to complement the *dl* mutant.

*Live imaging*

Live imaging of the Dorsal-Venus embryos was carried out using two-photon scanned light-sheet microscopy (Truong et al., 2011). Using a custom-built microscopy setup, 3-micron-thick light sheets were used to illuminate bidirectionally from two opposing sides of the embryo, creating an optical section that was perpendicular to the embryo's AP-axis (Figure 2A). The illumination light, derived from a femtosecond-pulsed laser (Chameleon UltraII, Coherent Inc.) was set at 960 nm to simultaneously induce fluorescence from Venus-labeled Dorsal proteins and RFP-labeled nuclei (H2A-RFP; Bloomington Stock Center) via two-photon excitation. To image the dynamic Dorsal-Venus nuclear gradient as shown in Figure 3, pre-nc10 embryos were mounted horizontally with heptane-glue on a coverglass with about 2/3 of the embryo's anterior body extending beyond the edge of the coverglass. The coverglass was mounted in a 25°C water-filled chamber and oriented so that the embryo's anterior end faced the detection optics. Imaging was conducted at a single focal plane, 150 microns from the embryo's anterior end; in 15-second time intervals; with 6 seconds of illumination/exposure time. The fluorescence from Dl-Venus and H2A-RFP were spectrally separated and imaged simultaneously onto neighboring regions of the recording camera (iXon-DU885, Andor Technology) with a spectral splitter (DV2, Photometrics).

*Image analysis*

In our analysis of Dorsal gradients, we took a similar approach as described in (Liberman et al., 2009), which is outlined here. First, the background was subtracted to set regions outside the embryo to black. Next, nuclei were detected (Figure 1C, lower right),

and the Dorsal fluorescence (Figure 1C, upper) was normalized to the nuclear data. For analysis of anti-sense RNA hybridized embryos, intensity as a function of arc distance was measured in an annular region around the periphery of the embryo. See also Supplemental Experimental Procedures for more details.

*Empirical fits of the Dorsal gradient*

The Dorsal nuclear gradient empirically conformed to Gaussian-like curves as previously explained (Liberman et al., 2009):

$$c_{dl}(x) \approx Ae^{-x^2/(2\sigma^2)} + B \quad (1)$$

Each gradient is thus represented spatially by three parameters (*A*, *B*, and σ). The first two parameters, *A* and *B*, describe the "amplitude" and "basal levels" of the gradient, respectively. The basal levels (*B*) can be thought of as the amount of nonzero Dorsal that is present in the dorsal-most nuclei. The amplitude (*A*) can be thought of as the amount of nuclear Dorsal present in the ventral-most nuclei greater than that found in the dorsal-most. σ is a measure of the spatial range of the gradient (gradient width).

To account for the gradual slope after 40% DV length, a correction to the Gaussian-like behavior (Equation 1) was made:

$$c_{dl}(x) \approx Ae^{-x^2/(2\sigma^2)} + B + M|x| \quad (2)$$

This final parameter, *M*, multiplies the absolute distance from the ventral midline, and denotes the value of the gradual slope found after the Gaussian term decays to zero. A linear function for the gradient tail was chosen over other one-parameter realizations because it is the simplest representation of a function with a nonzero slope, and can be

interpreted as a one-term Taylor expansion of the real functional form of the gradient tail. While this leads to a nonzero derivative at the dorsal midline, and thus is non-physical, the goal of this parameterization of the gradient tail is simply to distinguish between a flat and a non-flat tail, and is sufficient for that test (see Figure S4).

*Simulation of the wildtype Dorsal Gradient*

We used Equation 2 to simulate the wildtype Dorsal gradient in space and time (see Figure 7). We extracted $A(t)$ and $B(t)$ by averaging the data from three live embryo measurements and $\sigma = 0.14$ from wildtype fixed embryos. We chose $M(t) = -0.1A(t)$ to reflect the mean value of the normalized outer slope.

*Simulation of mRNA dynamics*

mRNA is described by the following equation:

$$\frac{d[mRNA]_i}{dt} = \frac{1}{\tau_i}(f_i - [mRNA]_i) \qquad (3)$$

where $i$ = *sna*, *vnd*, *sog*, or *zen*. $f_i$ is the mRNA production rate, modeled as a hard threshold function. For *sna*, this function is equal to 1 if $c(x,t) > \theta_{sna}$, and zero otherwise. In the case of *zen*, $f_{zen}$ is equal to 1 if $c(x,t) < \theta_{zen}$, and zero otherwise. For *sog* and *vnd*, $f_i$ is equal to 1 if both $c(x,t) > \theta_i$ and $[sna] < 0.5$, and zero otherwise. The input $c(x,t)$ is the simulated Dorsal nuclear gradient with 10% standard deviation Gaussian noise added. The production of mRNA only occurs during interphase. The parameter $\tau_i$ is the lifetime of mRNA species *i*.

The values of $\theta_i$ and $\tau_i$ were fitted to the data from Figure 4F. For further details on analysis of mRNA dynamics, see Supplemental Experimental Procedures.

*Calculations of precision*

To be able to read distinct Dorsal levels, nuclei that are spaced by a distance of $\Delta x$ must be able to measure Dorsal levels to within a relative error ($\Delta c/c$) given by the following equation:

$$difference\ between\ adjacent\ nuclei = \frac{\Delta c}{c} = \frac{1}{c}\frac{dc}{dx}\Delta x. \qquad (4)$$

During nc 14, the inter-nuclear distance is roughly 7 microns, and the equation for $c(x)$ is given by Equation 2, with $A = 1080$, $B = 520$, $\sigma = 0.14$, and $M = -89$. Here, $A$, $B$, and $M$ are in arbitrary units, and σ is in units relative to the length of the DV axis. These parameters are the average values for the fixed, nc 14 data set.

*Manual cross-sections of embryos*

For cross-section imaging, stained embryos in glycerol were manually cut with a 0.10mm blade under a dissecting microscope to remove the anterior and posterior ends, leaving a section 100-200 microns in width that corresponds to 150-200 microns from the embryo poles. These cross sections were then aligned on a glass slide and mounted in glycerol with a coverslip. Two pieces of double-sided tape were used as a spacer between the microscope slide and coverslip. Z-stacks of 15-20 microns were imaged using a Zeiss LSM 5 Pascal.

FIGURES AND LEGENDS

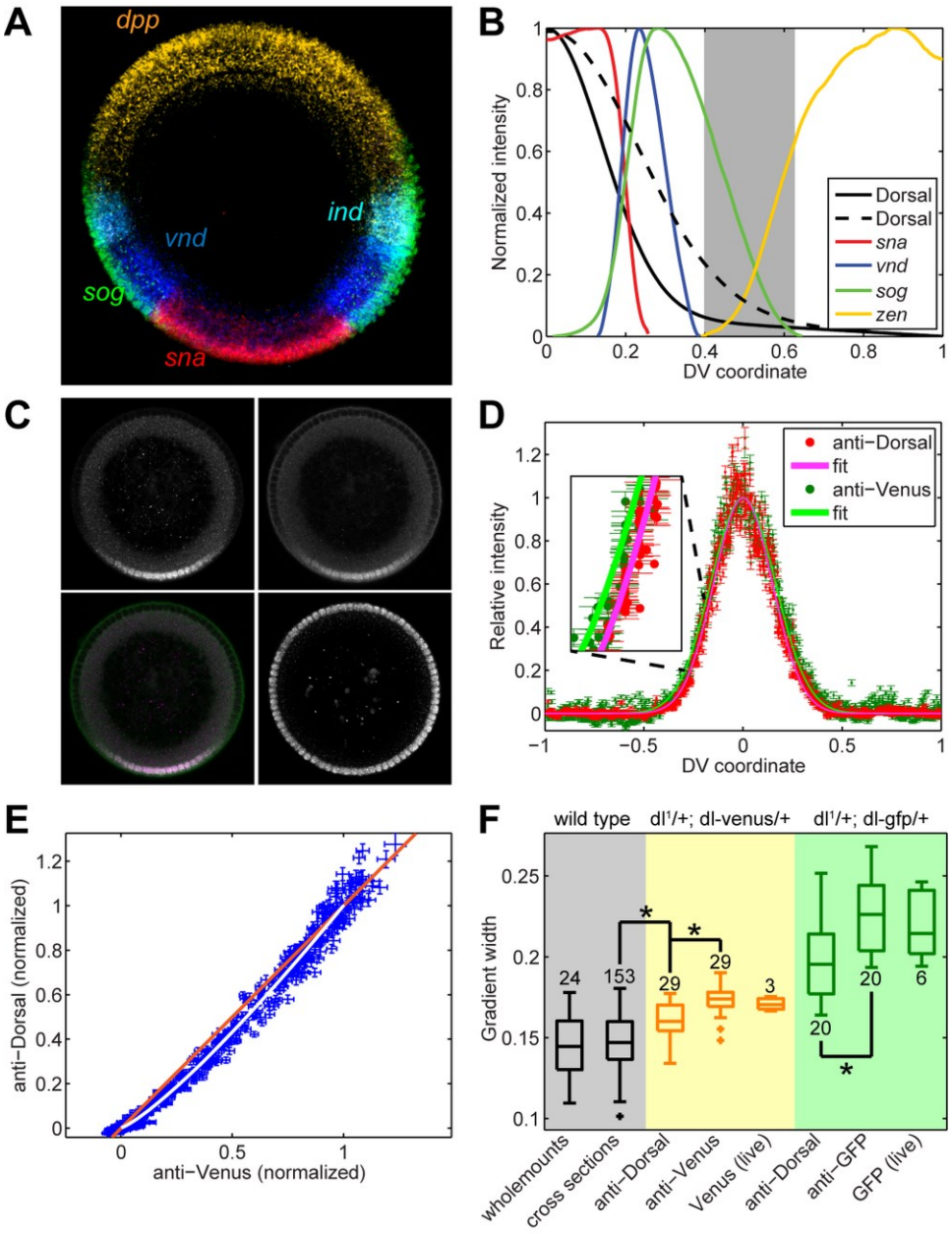



**Figure 1: Quantitative analysis of nuclear Dorsal levels and expression domains of dorsal-ventrally expressed genes.** (A) Cross section of late nc 14 embryo hybridized with *sna* (red), *vnd* (blue), *ind* (cyan), *sog* (green), and *dpp* (yellow) anti-sense RNA probes. (B)

Plot of domains of gene expression for Dorsal target genes. Data comes from averages of > 10 embryos for each gene. The solid black curve is average Dorsal gradient as measured from cross sections (see Figure 6A). The dotted Dorsal curve is from Bothma et al., 2010. The region where Dorsal's input is questionable appears in gray. (C) Antibody staining in a fixed Dorsal-Venus embryo cross section (clockwise from upper left): anti-Dorsal, anti-Venus, anti-histone H3, and merge between anti-Dorsal (magenta) and anti-Venus (green). (D) Quantification of fluorescent antibody staining from part C. Each dot corresponds to the intensity in a nucleus for anti-Dorsal (red) and anti-Venus (dark green). Errorbars denote the standard error of the intensity of the pixels in each nucleus (also in E). The two solid curves represent best-fit curves for anti-Venus intensity (green) and anti-Dorsal (magenta). This demonstrates that anti-Venus is slightly wider (see inset). (E) The normalized intensity of anti-Venus plotted against anti-Dorsal for each nucleus. Note that, in intermediate intensities, the curve falls below the 45 degree line (orange), indicating that anti-Venus is brighter on average than anti-Dorsal, and thus the gradient is wider. (F) Box plot of gradient widths ($\sigma$, see Equation 1 in Materials and Methods) for various antibody stainings, live imaging analysis, and maternal genetic backgrounds. Numbers indicate sample size. Plus signs indicate outliers. Asterisks indicate statistical significance (*p*-values $< 10^{-6}$). Wholemount data from Liberman et al., 2009. Embryo in (A) reprinted with permission from Reeves and Stathopoulos (2009), Graded Dorsal and differential gene regulation in the *Drosophila* embryo. *Perspectives on Generation and Interpretation of Morphogen Gradients*, eds Briscoe J, Lawrence P, Vincent J.-P. (Copyright 2009, Cold Spring Harbor Lab Press, Plainview, NY)."

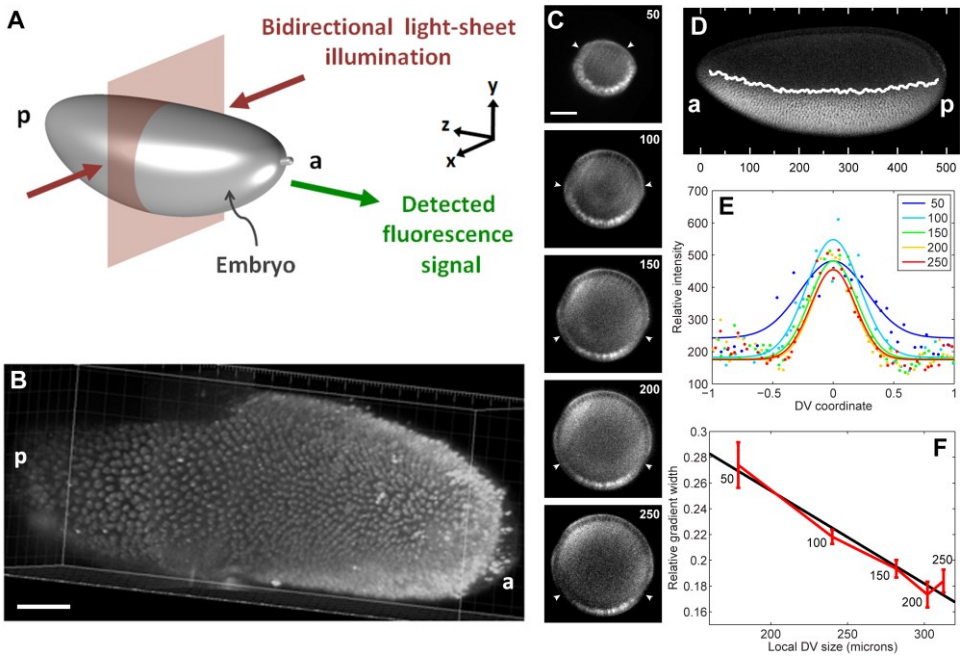



**Figure 2: Two-photon light-sheet microscopy reveals anterior-posterior modulation of the Dorsal gradient width.** (A) Schematic of the illumination and detection geometry used in live imaging of Dorsal-Venus embryos. (B) Three-dimensionally-rendered volume image stack of the nuclei-labeled H2A-RFP signal from a live embryo at nc 14 demonstrates that the nuclei-resolving resolution is achieved up to at least 250 microns from the anterior end (or ~50%) of the embryo. The optical distortions seen beyond 250 microns are due to the coverglass that supports the embryo. (C) Images of an embryo at 50-250 microns from the anterior pole. The white arrowheads denote the location where nuclear and cytoplasmic Dorsal-Venus become roughly equal in intensity. (D) Visual illustration of the gradient width as a function of AP position in a wholemount embryo fluorescently stained against anti-Dorsal. Hashmarks indicate distance (in microns) from the anterior pole. The white curve represents the approximate location where the nuclear and cytoplasmic intensity become equal. (E) Quantification of the Dorsal-Venus nuclear

gradients from part C. (F) The width of the gradients (σ, see Equation 1 in Materials and Methods) from C in units relative to the local DV size, plotted against the local DV size. Numbers next to the points denote distance from anterior pole. Errorbars indicate 68% confidence interval in computing σ. a: anterior, p: posterior. Scale bar, 50 microns.

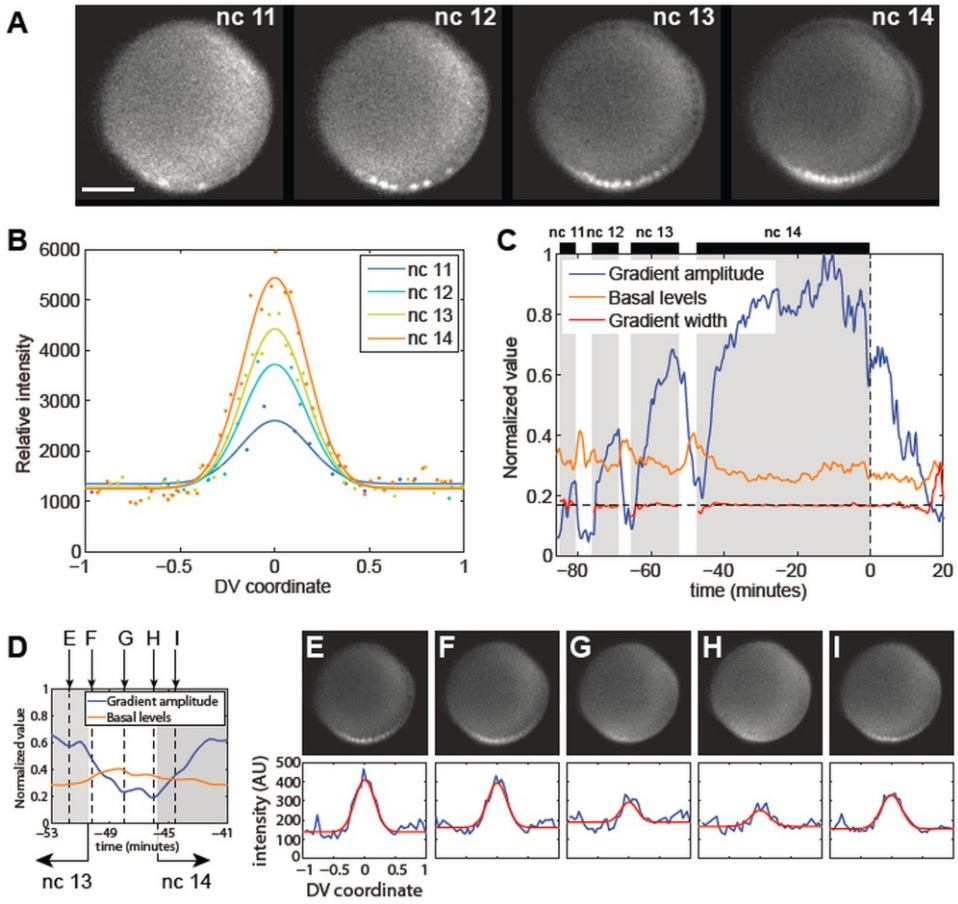



**Figure 3: The Dorsal-Venus nuclear gradient is dynamic, with increasing amplitude, decreasing basal levels, and constant width.** (A) Snapshots of the Dorsal-Venus nuclear gradient for nuclear cycles 11-14 imaged at 150 microns using two-photon light-sheet microscopy. These snapshots were taken at peak values of the gradient amplitude for each nuclear cycle. Scalebar, 50 microns. (B) Quantification of the Dorsal-Venus nuclear gradient from snapshots shown in part A. (C) Evolution of gradient amplitude (blue), basal levels (orange), and gradient width (red) from nuclear cycle 11 through gastrulation for the embryo shown in A. (D) Normalized gradient amplitude and basal levels from a single embryo zoomed-in on time points between 53 and 41 minutes before gastrulation. The mitosis between nc 13 and 14 interphases takes place between roughly 50 and 46 minutes

before gastrulation. The vertical dashed lines represent the time points depicted in the following panels. (E-I) Snapshots of the Dorsal-Venus gradient at the time points shown in D. The time points progress from the end of nc 13 interphase (E), the beginning of the following mitosis (F), the middle of mitosis (G), the end of mitosis (H), and the beginning of nc 14 interphase (I). Even in part H, a detectable ventral-to-dorsal gradient is present. Blue curves represent raw data. Red curves represent the Gaussian-like fit (Equation 1). See also Figure S1 and Movie S1.

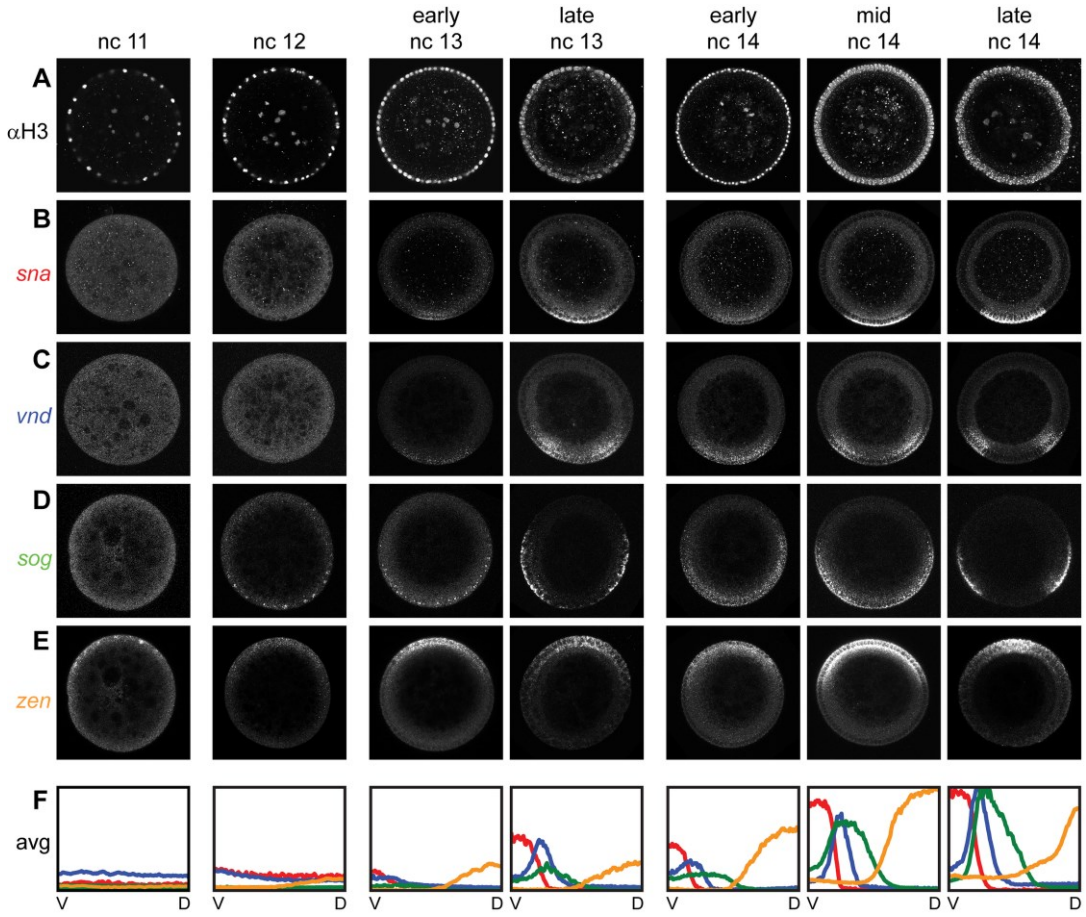



**Figure 4: Changes in mRNA patterns are identifiable both between and within nuclear cycles.** Wildtype embryos from double in situ and antibody fluorescent stainings were manually cross-sectioned and imaged. (A) Nuclei were labeled with α-Histone H3 to determine embryo stage. Expression of *sna* (B, Type I), *vnd* (C, Type II), *sog* (D, Type III+), and *zen* (E, Type III-) throughout nuclear cycles 11-14. Embryos are oriented with ventral side down. (F) Profiles of each gene (color-coded) reflect the expression averaged from 4-13 embryos at each nuclear cycle. Embryos were co-stained with *sna* and *vnd* or with *sog* and *zen*. D: dorsal, V: ventral. Brightness and contrast of embryo cross-sections have been adjusted for visual clarity. See Materials and Methods for analysis of raw data. See also Figure S2.

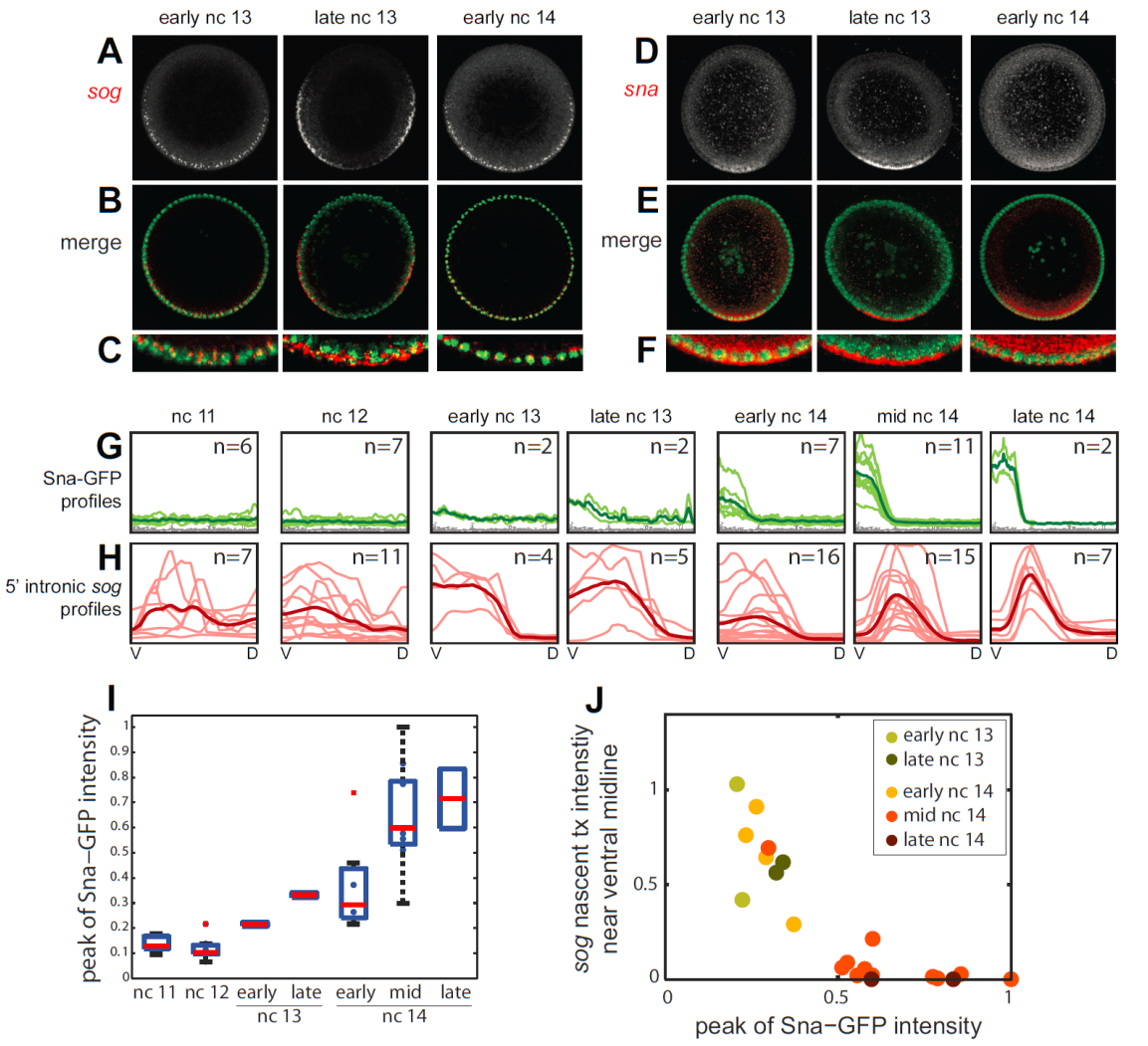



**Figure 5: Dynamics of nascent *sog* correlate with Sna protein.** (A) 19.5 micron Z-stack projections of *sog* mRNA. (B) Z-stack projection merge of *sog* mRNA (red) and nuclear Histone H3 staining (green). (C) Part B zoomed in to the area of *sog* expression. (D-F) Same as A-C, respectively, except with *sna*. (G) Profiles (light green) of Sna-GFP from all embryos analyzed (*n*) and the average curve (dark green). (H) Same as G except for intronic *sog*. Additional embryos stained with intronic *sog* but not GFP are included. (I) Box plot of peak Sna-GFP levels shows upward progression during nc 11-14. Red spots indicate outliers. (J) The intensity of intronic *sog* at the ventral midline is plotted against

the peak intensity of Sna-GFP. As the nuclear cycles progress, a decrease in intronic *sog* expression correlates with an increase in Sna-GFP. Only embryos co-stained with both intronic *sog* and Sna-GFP are included. See also Figure S3.

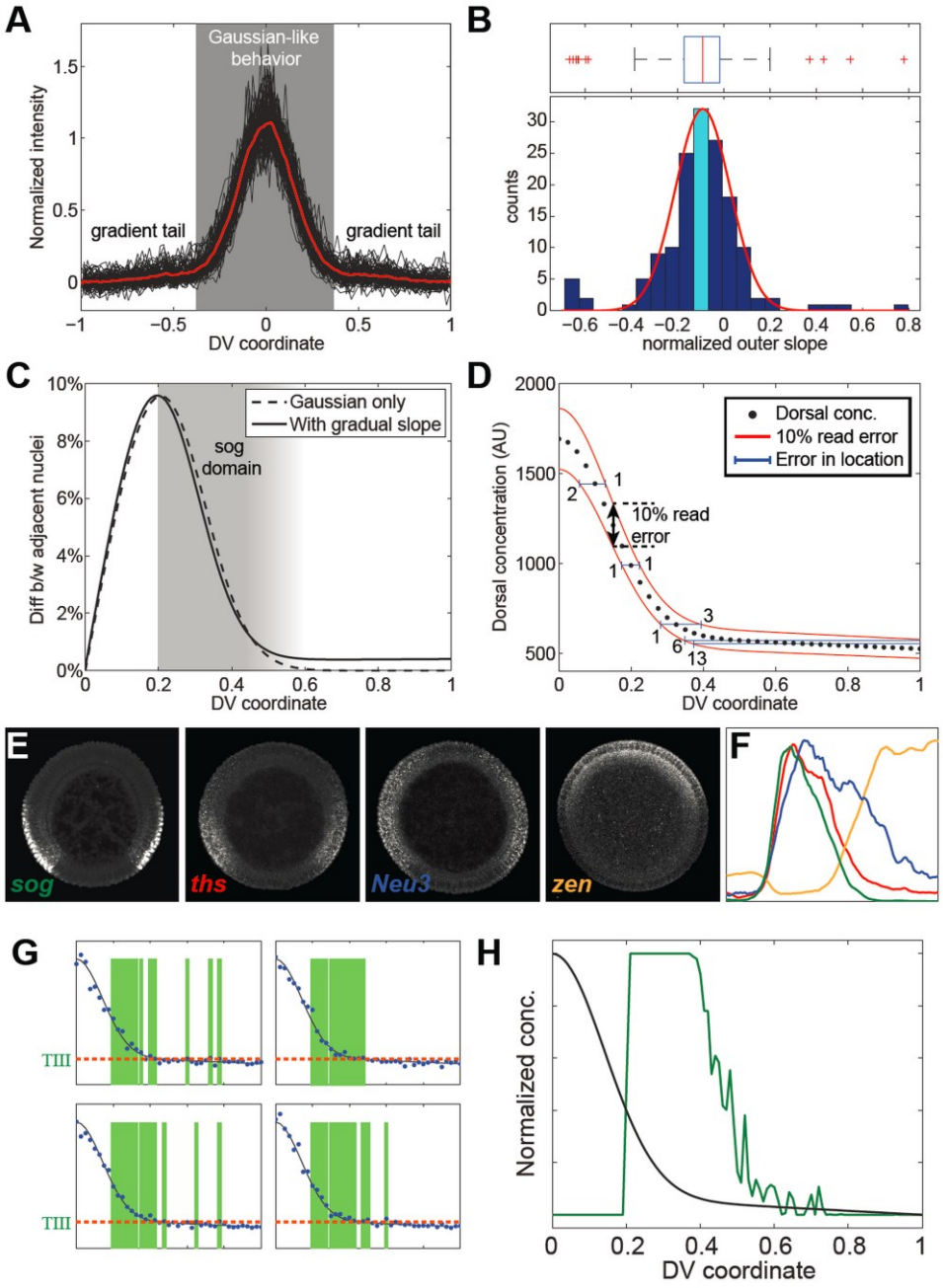



**Figure 6: A shallow gradient can deliver positional information with limited precision, supporting graded expression patterns.** (A) Normalized Dorsal nuclear gradients (black) for 153 fixed, manually cross-sectioned embryos. Average gradient in red. (B) Histogram of the normalized tail slope for all embryos from part A. The cyan histogram bar denotes

the 95% confidence interval of the mean. The box plot above the histogram depicts the bulk of the data falling within the box-and-whiskers, with a handful of extreme outliers both above and below the bulk of the data (red "+" signs). The red curve overlying the histogram depicts a normal distribution with the same mean and variance as the bulk of the data. (C) Plot of difference in Dorsal levels seen by adjacent nuclei versus DV coordinate. (D) Potential errors in gene expression boundary placement due to 10% stochastic fluctuations in Dorsal readout (red curves). From left to right, errorbars in $x$ denote the error for a gene presumptively placed at 10%, 20%, 33%, 50%, and 70% DV position. Numbers indicate rough numbers of nuclei. Each black dot represents a nucleus. 40 nuclei are plotted, in keeping with a typical nc 14 nuclear density. (E) Mature, Dorsal-dependent expression of Type III genes *sog*, *zen*, *ths*, and *Neu3*. *zen* pattern is shown before Dpp-dependent refinement occurs. (F) Profiles of genes shown in part E. Note differing locations of the dorsal boundaries and graded borders. (G) A noisy gradient tail may result in graded boundaries of Type III genes. Simulations of four instances of Type III+ gene activation as a result of reading out a noisy gradient (blue dots indicate the readout of each nucleus). In each case, gene expression is either active in a nucleus (green bar) or not, depending on whether the read Dorsal signal is above the threshold (red dotted line). Nuclei closer to the ventral side will be activated more often. (H) Final output of a Type III gene. This pattern is the result of time-averaging of the activation states of the nuclei (examples in G) and the basal levels decreasing within a nuclear cycle (example in 7E). The graded boundary of the Type III gene is determined by how long and how often nuclei read a signal above the threshold. See also Figure S4.

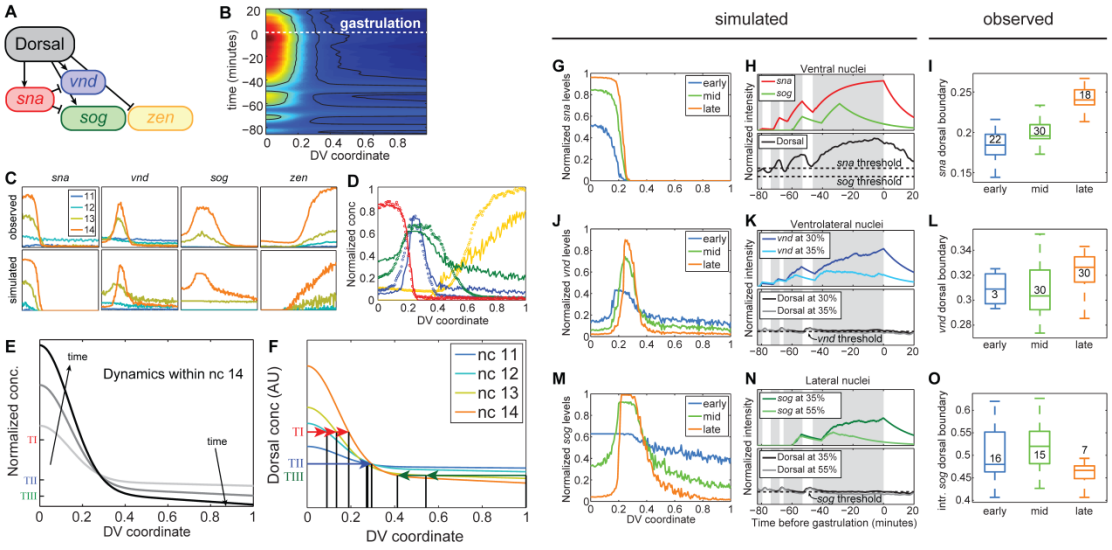



**Figure 7: Expression of DV genes in space and time correlates with dynamic nuclear Dorsal levels.** (A) Signaling network used in the model simulations. Arrows indicate activation; blunt arrows indicate repression. (B) Heatmap of the simulated Dorsal gradient. Black curves denote constant Dorsal concentration contours, corresponding to the thresholds chosen for *sna*, *vnd*, and *sog*/*zen*. (C) The observed (top row) and simulated (bottom) profiles of *sna*, *vnd*, *sog*, and *zen*. (D) The mid nc 14 Dorsal target genes *sna* (red), *vnd* (blue), *sog* (green), and *zen* (yellow). Circles denote averages of fluorescent in situ hybridization patterns from > 10 embryos and solid curves denote simulation results. (E) Simulation of dynamic Dorsal morphogen gradient. TI, II, and III were placed as in B, using the final gradient. (F) Simulations of threshold responses to Dorsal gradients from the nc 11-14 near the end of each nuclear cycle when the gradient amplitude is peaking. The horizontal lines correspond to the Type I (red), Type II (blue), and Type III (green) thresholds (contours of panel B). The locations these thresholds are crossed by the nc 11-14 Dorsal gradients are given by the vertical lines. (G) Simulated *sna* expression pattern for

early, mid, and late nc 14. (H) The amount of simulated Dorsal (black), *sog* (green), and *sna* (red) a ventral nucleus sees. The dashed horizontal lines correspond to the *sna* and *sog* thresholds. (I) Box plot of location of *sna* boundary in wild-type embryos, staged within nuclear cycle 14. Numbers indicate sample size. (J) Simulated *vnd* expression pattern for early, mid, and late nc 14. (K) The amount of simulated Dorsal (black, gray) and *vnd* (blue, cyan) seen over time by nuclei at 30% and 35% DV position, respectively. The dashed horizontal line corresponds to the *vnd* threshold. (L) Box plot of location of *vnd* boundary in wild-type embryos, staged within nuclear cycle 14. (M) Simulated *sog* expression pattern for early, mid, and late nc 14. (N) The amount of simulated Dorsal (black, gray) and *sog* (dark and light green) seen over time by nuclei at 35% and 55% DV position, respectively. The dashed horizontal line corresponds to the *sog* threshold. (O) Box plot of location of intronic *sog* boundary in wild-type embryos, staged within nuclear cycle 14. See also Figure S5.

CHAPTER III

# Image analysis and empirical modeling of gene and protein expression

This chapter was written with Alphan Altinok, Angelike Stathopoulos, and Gregory T. Reeves, and published in *Methods,* 62(1):68-78 in 2013.

ABSTRACT

Protein gradients and gene expression patterns are major determinants in the differentiation and fate map of the developing embryo. Here we discuss computational methods to quantitatively measure the positions of gene expression domains and the gradients of protein expression along the dorsal-ventral axis in the *Drosophila* embryo. Our methodology involves three layers of data. The first layer, or the primary data, consists of z-stack confocal images of embryos processed by in situ hybridization and/or antibody stainings. The secondary data are relationships between location, usually an $x$-axis coordinate, and fluorescent intensity of gene or protein detection. Tertiary data comprise the optimal parameters that arise from fits of the secondary data to empirical models. The tertiary data are useful to distill large datasets of imaged embryos down to a tractable number of conceptually useful parameters. This analysis allows us to detect subtle phenotypes and is adaptable to any set of genes or proteins with a canonical pattern. For example, we show how insights into the Dorsal transcription factor protein gradient and its target gene *ventral-neuroblasts defective* (*vnd*) were obtained using such quantitative approaches.

# 1. INTRODUCTION

In a developing animal, the distributions of signaling proteins, termed "morphogens", dictate the patterning of gene expression within developing tissues in a concentration-dependent fashion (Wolpert, 1969). High morphogen concentrations drive the expression of one set of genes, while low concentrations a different set. In this manner, a single protein, distributed in a spatial gradient, can be responsible for the gross patterning of an entire tissue. Therefore, to model cell-cell signaling and gene expression patterns in development, quantitative measurements of protein and RNA distribution within the embryo are necessary. Fluorescent experimental techniques, such as fluorescent antibody staining and fluorescent in situ hybridization, are sufficiently quantitative for such measurements (Goentoro et al., 2006; He et al., 2008; Liberman and Stathopoulos, 2009; Luengo Hendriks et al., 2006). Yet, to extract these measurements from the fluorescent image data, and in order to make meaningful comparisons across several embryos and sets of embryos, an image analysis protocol is needed (Ay et al., 2008; Fowlkes et al., 2008; Jaeger et al., 2004; Surkova et al., 2008).

Here we present a protocol to analyze image data consisting of three types of molecular species: (1) mRNA/non-nuclear proteins, (2) nuclear proteins, and (3) nascent transcripts (via intronic probes; sometimes called "nuclear dots"). The goal of this method is to distill fluorescent images down to a set of meaningful parameters that characterize the protein and mRNA distributions. To this end, we identify three levels of data: the fluorescent images are the primary data, secondary data are the relationships between position on the embryo and fluorescent intensity, and the tertiary data are parameters that

arise from fitting our secondary data to empirical models of protein and mRNA localization.

The final, data-fitting step requires the development of accurate models of protein and gene expression, as well as numerical techniques for minimization of error. However, using the protocol of data fitting to empirical models has several advantages over more simplistic approaches. First, it involves the use of the entire secondary data set, rather than small sets of distinct points. Second, because it uses the entire data set, it is robust to noisy secondary data. Third, it allows for a systematic analysis of variance in the predicted parameters. Because of these advantages, we can be confident in our ability to detect very subtle phenotypes that otherwise are difficult to discern, either by eye or through comparison between sets of secondary data.

As a concrete example, we focus on the transcription factor Dorsal (dl), which acts as a morphogen in the early *Drosophila* embryo (rev. in Moussian and Roth, 2005). dl nuclear localization is regulated spatially in the *Drosophila* embryo, in a gradient, such that high nuclear levels are present in ventral regions and very low nuclear levels are present in dorsal regions (Figure 1A). High levels of nuclear Dorsal support expression of genes such as *snail* (Figure 1B,C) and *twist*, and intermediate levels of nuclear Dorsal support expression of genes such as *ventral neuroblast defective* (*vnd*; Figure 1A,C) and *short gastrulation* (*sog*; Figure 1C). Dorsal can also function as a repressor of transcription and in this function limit the expression of some genes, such as *decapentaplegic* (*dpp*) and *zerknüllt* (*zen*), to dorsal regions (rev. in Reeves and Stathopoulos, 2009; Stathopoulos and Levine, 2005).

In this paper, we describe the steps in the image analysis protocol and give details of the calculations involved in the data fitting procedures. We demonstrate two cases in which our analysis protocol has allowed us to detect subtle, yet statistically significant, phenotypes in the dl patterning system. In the first case, we detect the slightly longer decay length in the nuclear gradient of a Venus-tagged version of dl, as compared to the nuclear gradient of wildtype dl (Reeves et al., 2012). The second case consists of measuring the subtle perturbation in placement of the *vnd* dorsal boundary in embryos that exhibit wider gradients. We also present an example of image analysis in a system other than a cross section of the Drosophila embryo.

# 2. MATERIALS AND METHODS

Here we briefly present steps for experimental preparation of *Drosophila* embryo cross-sections and the mounting and imaging conditions required by our method. More details on the image analysis procedure can also be found in the supplementary material.

## 2.1 Collection, fixation and in situ hybridization of embryos

Fly stocks used in Figure 8 include *dl-venus* and *dl-gfp* constructs and are further described in (Reeves et al., 2012). Embryos in Figure 9 come from the stock $dl^{1}$*/CyO; dl-gfp, H2A-rfp/TM3*, and embryos in Figure 10 are F10 mutants (Huang et al., 1997). Fluorescent in situ of *vnd*, *ind*, *sna*, and *sog* genes using antisense RNA probes, and antibody stainings of Dorsal (DSHB), histone H3 (Abcam), and GFP (Rockland) were performed using standard protocols. These steps proceeded according to published protocols (Kosman et al., 2004). When detecting protein distributions, the proteinase K treatment is skipped. After completion of the last wash, embryos are stored in 70% glycerol at -20°C in the dark.

## 2.2 Embryo manipulation

After the fluorescent in situ hybridization/antibody staining protocol is performed, embryos are transferred to a viewing dish with 70% glycerol. Using a brush, the appropriate stage is selected by morphology under a stereomicroscope and placed on a glass slide. Generally, ~10 embryos are sectioned at a time. This prevents over-drying and reduces exposure to light. A small amount of 70% glycerol may need to be added onto the

slide to prevent embryos from desiccating. However, too much glycerol will cause difficulty in maneuvering the embryos. These embryos are then manually cross-sectioned with a 0.10 mm blade and mounted upright using a hair loop. Two pieces of double-sided tape are used to prevent pressure on and damage to the sections. A coverslip is placed on top, and 70% glycerol is pipetted between the slide and coverslip.

**2.3 Confocal imaging**

Cross-sectioned embryos were imaged with a 40x 1.3 NA oil objective; pinhole of 2.29 AU; pixel time of 3.20 μs. 15-20 slices with 1.3 μm-thickness containing the middle region of the embryo cylinder were acquired. For the success of the image analysis, it is crucial to have only the desired embryo in the image, and no other fluorescent materials, such as other embryos or dust. It is also imperative to have the entire embryo within the image for all z-slices; leave a comfortable padding of black space around the embryo. See Figure 1A for an example.

**2.4 Image analysis**

There are five steps in the image analysis procedure, as outlined below. These procedures were developed and applied in recent studies (Liberman and Stathopoulos, 2009; Reeves et al., 2012). In those cases, we will provide references for additional background and a brief discussion.

*2.4.1 Detecting the border of the embryo*

This procedure was first introduced in (Liberman and Stathopoulos, 2009) and utilized more recently in (Reeves et al., 2012). First, the geographical center of the image is assumed to reside inside the embryo. From this location, the image is divided into 60 slices in the azimuthal angle (i.e., $\theta$; see Figure 2A). The intensity of the image as a function of $r$ (the distance from the center of the image) is found for each slice in $\theta$ (Figure 2B). The presumed location of the boundary of the embryo is where the intensity drops rapidly (red dot in Figure 2B). This point in $(r, \theta)$ coordinates is then transformed back into a pixel location (i.e., $xy$ coordinates), resulting in a relatively tight border around the embryo (Figure 2C).

*2.4.2 Calculation of average intensities*

The next step is to detect average intensities around the periphery of the embryo, and is described briefly in (Reeves et al., 2012), and in more detail here. The average intensities serve as measures of gene expression in color channels corresponding to mRNA probes or averaged values of other molecular species for which nuclear localization is not an important factor. This analysis proceeds in three steps. First, from the 60 points on the embryo periphery, a much denser outer ring of 300 points is interpolated (Figure 3A). For each point $i$ on the outer ring, the outer ring points $i, i+1$ and two corresponding points, roughly 20 microns in from the periphery, form a quadrilateral (Figure 3B). Third, the average intensity inside this quadrilateral, $t_i$, is taken as the intensity value of this color channel at point $s_i$. The result of this analysis step is a vector of pseudo-arclength, $s$, and a

corresponding array, $t$, which contains the intensities of each color channel as a function of $s$. For example, a plot of $t$ vs. $s$ will yield peaks of gene expression if the color channels in the image are fluorescent representations of mRNA probe hybridization (Figure 3C).

*2.4.3 Locating nuclei*

In the case of images with nuclear stains (such as antibodies against histone proteins, or other dyes like TOPRO or DAPI), the analysis program can locate the nuclei (Reeves et al., 2012). We briefly describe this analysis here, which proceeds in several steps. First, using the 60 points on the border of the embryo, as described in Section 2.4.1, the outer periphery of the embryo cross section (up to 20 microns deep) is "unrolled" into a long strip (Figure 4A), which we transform into a binary nuclear mask (Figure 4B). To accomplish this, the strip of nuclei is averaged in the radial (i.e., the apical-basal) direction to yield a one-dimensional representation of the nuclei (Figure 4C). Next, a watershed algorithm is used to determine the 1D locations of the cytoplasmic regions between nuclei. This allows us to put boundaries between the nuclei. These 1D locations are then mapped back onto the original unrolled strip to define rectangular regions, and inside of each rectangle there is exactly one nucleus.

Within each rectangle, the nucleus is segmented using a hard threshold, with the threshold level chosen using a best-fit protocol (Otsu, 1979). This local thresholding results in the ability to segment each individual nucleus from the surrounding cytoplasm (Figure 4B). Defining each rectangle is necessary to (1) differentiate between nuclei that

are touching or almost touching, as each nucleus is given a distinct numerical label, and (2) avoid the problems associated with using a global threshold algorithm on the entire image.

Afterward, the pixels corresponding to each nucleus in the unrolled strip are mapped back to the original 2D image of the embryo (Figure 4D). As this is a non-linear transformation, and not a one-to-one mapping, sometimes this results in solid nuclei with a handful of black pixels. These pixels are then "filled-in" with the numerical label corresponding to the nucleus they reside in (see insets in Figure 4D). This ensures the nuclear mask does not have missing pixels. Note this hole-filling step does not distort the data in any way; it is simply for complete labeling of individual nuclei.

The 2D image is then morphologically opened with a disk-shaped structuring element 5 microns in diameter. This removes the spurs and feathers from the nuclei, resulting in smooth nuclei (Figure 4D). At this point, the nuclear mask is complete, where the pixels in nucleus $i$ have a value of $i$, so that even nuclei that touch after transforming back into the 2D image are distinguishable (Figure 4E). All other pixels are black.

In some cases, the quality of the primary data is not high enough to reliably distinguish nuclei from their neighbors, resulting in ill-defined boundaries between neighboring nuclei. In those cases, the program therefore lumps together multiple nuclei (arrowheads in Figure 4E).

The important parameters associated with each nucleus are the pixel list (so the location of the nucleus can be mapped into the other color channels), the centroid (which

acts as single coordinate for the whole nucleus) and the nuclear intensity (which will act as a normalization factor for the other color channels; see Section 3.1).

It is important to point out that this method of detection of nuclei works very well in nuclear cycles (nc) 13 and 14 of *Drosophila* development, due to the relatively close-packing of the nuclei. Earlier nuclear cycles, such as nc 11 and 12, have a much lower density of nuclei, making it difficult to avoid false positives. We previously used manual nuclear detection to overcome this difficulty (Reeves et al., 2012); however, manual detection of nuclei is not discussed in this set of code.

### *2.4.4 Intensity of nuclear proteins*

Once the nuclear mask is determined, if the given embryo has been immunostained for one or more nuclear-localized proteins, the fluorescent intensity of the channels corresponding to this (these) nuclear protein(s) is considered. For each nucleus identified in the nuclear mask, the corresponding pixels in the nuclear protein channel are found, and the average intensity of those pixels is counted to be the intensity of that nuclear protein for that nucleus. Standard errors of the mean for each of these measurements are also calculated. Coupled with the data for the centroid of each nucleus, these measurements give us a relationship between nuclear protein intensity and location on the periphery of the embryo.

### *2.4.5 Measuring nuclear dots*

If the given embryo has been treated with anti-sense RNA probes made against intronic portions of genes (i.e., intronic probes), then nascent transcripts within nuclei

("nuclear dots") can be measured. For each nucleus identified in the nuclear mask (Figure 5A), the highest intensity pixel in the color channel corresponding to the intronic probe is counted as the center of the nuclear dot (Figure 5B,C). Against the possibility this high intensity pixel is a random high intensity photon as measured by the photomultiplier tube, the median intensity in a 5-by-5 pixel neighborhood centered on the high intensity pixel (red box in Figure 5B inset) is chosen to be the strength of the nuclear dot. Coupled with the data for the centroid of each nucleus, these measurements give us a relationship between intronic dot intensity and location on the periphery of the embryo. This relationship reveals a "salt-and-pepper" pattern, in which a wide range of values occurs at any given location in the domain of gene expression (blue dots in Figure 5D). Furthermore, the same dot is likely represented as multiple datapoints, because it may be present in multiple slices of the z-stack.

In order to fit the pattern of nascent transcription to a smooth profile, we exchange the multi-valued relationship between nuclear dot intensity and location for a single-valued, smooth curve (red curve in Figure 5D). We do this in a similar manner to what was described previously (Reeves et al., 2012). First, the locations of the nuclei are binned into a mesh from minus one to one with 300 points. Next, for every bin that contains nuclei, the top 5 nuclear intensities are averaged together. This averaged value becomes the value of our raw curve at that location. For bins that have fewer than 5 nuclei, all nuclei are averaged together.

If bin $i$ has zero nuclei, a value at location $i$ must be chosen in order to maintain a smooth curve. The program searches bin $i-1$ for a value. If no value is found, the

program searches bin $i-2$. This continues until a value is found to the left of bin $i$. Next, the program searches in a like manner to the right of bin $i$. After finding the closest values to the right and to the left of bin $i$, the value at bin $i$ is the average between these two values.

Finally, the resulting curve is smoothed with a sliding window of five points, resulting in a curve similar to the red curve seen in Figure 5D.

# 3. CALCULATION

After the image analysis procedures, the primary data has been transformed into the secondary data, consisting of relationships between intensities of the fluorescent readouts of the molecular species probed and location on the periphery of the embryo. There are two aspects to further calculate to fit our secondary data to empirical models of protein and gene expression, resulting in the fitted parameters that comprise our tertiary data. The first is normalization with respect to nuclear intensity, and the second is data fitting. More details can also be found in the supplementary material.

## 3.1 Normalization with respect to nuclear intensity

Due to uneven illumination and loss of light collection with z-depth, we have found it useful to normalize the nuclear protein intensities with respect to intensities of the nuclear stain (Liberman and Stathopoulos, 2009). To ground this normalization in theory, we make the following assumptions. First, we assume the relationship between measured intensity, $I_i$, of a nuclear protein in nucleus $i$ and the real nuclear concentration of the protein, $c_i$, is as follows:

$$I_i = k_i(\Phi_1 c_i + \Phi_2)$$

Here, the factor $k_i$ depends on the light path from the objective to nucleus $i$, and the constants $\Phi_1$ and $\Phi_2$ depend on experiment-wide factors, such as microscope settings, the concentration of antibody used, or the non-specific affinity of the antibody for embryonic tissue. In other words, when uneven illumination is taken into account, the fluorescent intensity is proportional to the protein concentration with an additive background constant.

In a similar vein, we assume the intensity, $N_i$, of the nuclear stain in nucleus $i$ is related to the actual concentration, $n_i$, of the molecular species by:

$$N_i = k_i(\nu_1 n_i + \nu_2)$$

We also assume the concentration of the nuclear species is the same in each nucleus throughout the embryo, meaning the term $\nu_1 n_i + \nu_2$ can be simply represented by a constant $H$. This implies

$$N_i = k_i H$$

Therefore, to eliminate the unknown dependence of $I_i$ on the lightpath-dependent factor, $k_i$, we normalize $I_i$ by $N_i$:

$$t_i = \frac{I_i}{N_i} = \frac{k_i(\Phi_1 c_i + \Phi_2)}{k_i H} = \phi_1 c_i + \phi_2,$$

where the new constants $\phi_1$ and $\phi_2$ are the old constants ($\Phi_1$, $\Phi_2$) divided by $H$. Thus, this normalized intensity, $t_i$, is simply proportional to the concentration, $c_i$, up to an additive background constant.

After normalizing each nuclear protein intensity with respect to the corresponding nuclear stain intensity, the value of $t$ is typically close to one. To return $t$ to a scale similar to its original intensity, we multiply by the mean intensity of all nuclei.

### 3.2 Data fitting

We next fit our secondary data to empirical curves. Doing so will result in a set of parameters that describe each of our secondary data relationships. We do this for mRNA, nascent transcripts, and nuclear-localized proteins.

*3.2.1 mRNA*

We fit our data of mRNA expression to "canonical" peaks of gene expression (Liberman and Stathopoulos, 2009; Reeves et al., 2012; see also Figure 6A). In so doing, we can measure, in a systematic way, the dorsal and ventral boundaries of gene expression. Each canonical gene expression profile was obtained by manual alignment of 10 or more instances of gene expression. Once this profile is found, it can be manipulated by translation, stretching/shrinking, addition, and multiplication, in order to fit to a measured gene expression profile. We have found that virtually every measured profile of gene expression can be fit to the canonical profile this way.

Consider a canonical gene expression profile, $y_c(x)$, with a peak located at $x_c$, where $x$ is the coordinate along the DV axis of the embryo (Figure 6A). Each measured gene expression profile, $y(x)$, can be fit to $y_c(x)$ in the following way:

$$y(x) \approx \alpha y_c\left(\frac{(x - x_c) - x_0}{\delta}\right) + \beta$$

In this equation, the two important parameters are the peak location, $x_0$, and the stretching factor, $\delta$. These two parameters together dictate how the canonical gene expression profile is changed in space to accommodate the measured profile (Figure 6B,C). From these two parameters, we can calculate the dorsal and ventral borders of our measured gene expression profile.

As an example, consider the *vnd* profile of the embryo in Figure 1A (see Figure 6D). After background subtracting this profile, with the appropriately-sized structuring element (Reeves et al., 2012), and plotting the two halves of the embryo on top of each other, we arrive at Figure 6E (circles; cyan corresponds to the right side of the embryo, and

magenta to the left side). The fitted canonical profiles are plotted as solid curves, with blue corresponding to the cyan data points, and red to the magenta. These fits resulted in the parameters $\delta_R = 0.928$, $x_{0,R} = 0.272$ and $\delta_L = 0.957$, $x_{0,L} = 0.263$ (where $R$ denotes the right side of the embryo, and $L$ denotes the left).

The canonical gene expression profile, $y_c(x)$, has a dorsal and ventral border associated with it, defined as the location in which the peak drops to half-maximal intensity (Figure 6A). If these two locations are denoted as $x_{d,c}$ and $x_{v,c}$, respectively, then $x_d$ and $x_v$ (the dorsal and ventral borders of the measured gene expression profile) are given by:

$$x_d = (x_{d,c} - x_c)\delta + x_0$$

$$x_v = (x_{v,c} - x_c)\delta + x_0$$

Therefore, each gene expression profile for each embryo has the parameters $\delta$, $x_0$, and from these, the dorsal and ventral borders of gene expression can be found quantitatively. Returning to our example of *vnd*, the canonical borders of gene expression for *vnd* are $x_{d,c} = 0.572$ and $x_{v,c} = 0.451$, with $x_c = 0.5$. Using these numbers in the example, $x_{d,R} = 0.339$ and $x_{d,L} = 0.331$; $x_{v,R} = 0.227$ and $x_{v,L} = 0.216$. However, we report the results of the dorsal and ventral borders of gene expression as the average of the borders calculated from both sides, meaning $x_d = 0.335$ and $x_v = 0.221$.

This approach has advantages over simply finding the location where the measured gene expression profile crosses half-maximal intensity for three reasons. First, if the measured gene expression profiles are noisy, then both the maximal intensity and the locations of half-maximal intensity cannot be found reliably. Second, finding the borders directly uses only a handful of data points from the measured gene expression profile, and

thus may be inaccurate. Finally, using only a couple of data points does not allow the calculation of confidence intervals on the parameter estimates. Using the fitting procedure described here overcomes these problems.

The two parameters $\alpha$ and $\beta$ refer to the max intensity of the peak and the background levels of image intensity, respectively. In general, neither of these fitted parameters informs us about the gene expression pattern, given that image intensity and background can be altered by microscope settings and slight differences in experimental procedure. However, if these factors are controlled for, then it is possible to perform a semi-quantitative analysis on the strength of gene expression, in which fluorescent intensities of either protein or gene expression can be compared embryo-to-embryo (Reeves et al., 2012). (See Section 5 Discussion for more details.)

Our analysis package uses the Matlab function "fit" to find these parameters. Additionally, the function returns confidence estimates on the fitted parameters, $\alpha, \beta, \delta$, and $x_0$. We propagate these confidence intervals from $\delta$ and $\mu$ to $x_d$ and $x_v$ by the following formulae:

$$dx_d = \sqrt{\left(x_{d,c} - x_c\right)^2 (d\delta)^2 + (dx_0)^2}$$

$$dx_v = \sqrt{\left(x_{v,c} - x_c\right)^2 (d\delta)^2 + (dx_0)^2}$$

Here, the terms $d\delta$, and $dx_0$ represent the radii of the 68% confidence intervals on these parameters, and can be thought of as the magnitude of one standard deviation. The interpretation of $dx_d$ and $dx_v$ are thus the magnitude of one standard deviation in these borders. These uncertainty measurements are typically on the order of one tenth of one

percent of the DV axis, which is less than one nucleus wide. For our example embryo, $dx_d = 0.0022$ and $dx_v = 0.0017$.

Profiles of nascent transcripts and nuclear proteins can also be fit using this procedure. To fit profiles of nascent transcripts, the smoothed profile (red curve in Figure 5D) is treated as the measured gene expression profile, $y(x)$. For nuclear proteins, see below (Section 3.2.2).

One thing to point out is this fitting procedure requires the ventral midline of the embryo to have previously been identified. This can be done either manually (see Example 3, Section S1.8.7 in the supplementary material), through the fitting of another molecular species that allows for unambiguous identification of the midline (e.g., the dl nuclear gradient; see Section 3.2.2), or through a rules-based procedure. Our current formulation employs the rules-based procedure, in which the program looks for certain kinds of peak maxima depending on which genes the user supplies. Once these peak maxima are found, the location of the midline is inferred. For example, if the gene is *vnd*, which is expressed in a symmetric, two-stripe, ventral-lateral pattern, the program locates the peak maxima and places the midline directly in between.

Another feature of this fitting procedure is that it can detect and resolve multiple gene expression profiles in the same color channel, as long as the two profiles are sufficiently separated in space. This is illustrated in Figure 1C and also in Figure 10. This feature can be extended to genes that have multiple domains of expression, such as *rho* and *hb* (see Section S.1.8.9 for an example).

*3.2.2 Nuclear proteins*

We originally developed this protocol to analyze the dl nuclear gradient. According to our previous analysis, the dl gradient can be empirically fit to a Gaussian-like function plus a slowly declining tail (Liberman and Stathopoulos, 2009; Reeves et al., 2012):

$$c_{dl}(x) \approx A \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) + B + M|x| \qquad (1)$$

Using this empirical model, the amplitude of the peak in the dl gradient is $A$, the gradient "basal levels" (i.e., the value where the gradient ceases to decay according to $x^2$) is $B$, the length scale of the Gaussian-like behavior is $\sigma$, the location of the ventral midline is $\mu$, and the slope of the shallowly declining gradient tail is $M$.

As an example, consider the dl nuclear gradient of the embryo in Figure 1A. After the image analysis step, we can plot the secondary data (the value of $t$ for each nucleus) vs. nuclear location (blue dots with errorbars in Figure 7). Fitting these secondary data to Equation 1 (see also black curve in Figure 7), we find the value of the Gaussian length parameter, $\sigma$, to be 0.15, and the normalized slope of the gradient tail, $M/A$ , to be -0.17.

Despite the fact that Equation 1 appears quite specific to dl, this can still be useful in other situations. For example, it is plausible that the pattern of pMad in the early embryo conforms well to Equation 1. Even when the nuclear protein of interest does not conform to Equation 1, as was the case with Sna-GFP [13], fitting the nuclear protein profile to Equation 1 can often reliably find the midline, amplitude, and basal levels of the profile. The parameter σ will also be correlated to width of the profile, but a strict definition of σ will be lost.

However, nuclear proteins that do not conform to Equation 1 can alternatively be fit to a canonical profile. As this is the more general case, we have written the workflow of our program to fit any nuclear protein not labeled ‘dl’ to a canonical profile (See S1.2).

As in fitting mRNA profiles, we use the Matlab function “fit” to find these parameters as well as their 68% confidence intervals.

# 4. RESULTS

We use the dl morphogen system in nuclear cycle 14 (2-3 hour old) *Drosophila* embryos as a test case for our analysis and data-fitting program. Here we present two example scenarios that demonstrate our protocol can detect very subtle phenotypes not easily distinguished by eye. We also present a third example scenario that helps describe the generalizability of the program to alternative geometries.

## 4.1 Expanded dl nuclear gradient in embryos carrying a copy of Venus-tagged dl

The wildtype dl nuclear gradient conforms to a Gaussian-like decay with sloping tails (Equation 1; Liberman and Stathopoulos, 2009; Reeves et al., 2012). In previous work, we measured the spatial extent of the dl gradient, which is an important parameter in determining the ability of dl to pattern the DV axis. For wildtype embryos, we have found the width of the dl gradient to be roughly $\sigma_{wt} = 0.14 \pm 0.01$ (Figure 8D; Liberman and Stathopoulos, 2009; Reeves et al., 2012). Using our protocol, we can detect a subtle, systematic difference between the width of the wildtype dl nuclear gradient and width in embryos with perturbed dl gradients. Here, we present a case in which embryos from mothers with one null allele of endogenous *dl* and one copy of YFP *venus*-tagged *dl* (hereafter referred to as *dl-venus* embryos) have a statistically detectable increase in gradient width.

We detected the dl nuclear gradient in fixed *dl-venus* embryos both by using an antibody directed against dl (upper left panel in Figure 8A) and an antibody directed against GFP (which also recognizes Venus; upper right panel in Figure 8A). Judging from

the similarity between the two images, we could make a case these two gradients are identical (for comparison, see merged image in the lower left panel in Figure 8A). However, the secondary data obtained from our image analysis protocol shows the gradient detected by anti-GFP is wider than that detected by anti-dl (Figure 8B,C). Using our empirical model of the dl gradient, we measured the widths of these two gradients to be $\sigma_V^{dl} = 0.16 \pm 0.01$ for detection by anti-dl, and $\sigma_V = 0.17 \pm 0.01$ for detection of Venus by anti-GFP (mean ± std dev; Figure 8D). This difference between these distributions in widths across the set of 29 embryos is statistically significant, using the t-test for correlated samples (p-value < $10^{-6}$; Reeves et al., 2012).

We also found the gradient detected by anti-dl in *dl-venus* embryos is wider than the dl nuclear gradient in wildtype embryos. In summary, $\sigma_{wt} < \sigma_V^{dl} < \sigma_V$, implying the extent of the actual spatial gradient of dl-Venus in the *dl-venus* embryos is wider than the gradient of endogenous dl (compare $\sigma_V = 0.17$ to $\sigma_{wt} = 0.14$) in these same embryos. Since the anti-dl antibody will recognize both endogenous dl and dl-Venus, the anti-dl measurement in these embryos ($\sigma_V^{dl} = 0.16$) is some intermediate value. Furthermore, the measurement of the dl-Venus gradient using anti-GFP is upheld by measurements of Venus fluorescence in live *dl-venus* embryos (Figure 8D).

This same trend is also seen in embryos from mothers with one null allele of endogenous *dl* and one copy of *gfp*-tagged *dl* (hereafter referred to as *dl-gfp* embryos; see Figure 9). The gradient width as detected by anti-dl ($\sigma_G^{dl} = 0.20 \pm 0.02$) is not as wide as that detected by anti-GFP ($\sigma_G = 0.23 \pm 0.02$), yet both are wider than the wildtype

gradient. Live studies of these embryos also show the anti-GFP measurement to be accurate (Figure 8D).

### 4.2 Shifted *vnd* dorsal border in embryos with expanded dl gradients

We have also asked if the expanded width of the dl gradient in *dl-gfp* embryos has an effect on gene expression. We immunostained *dl-gfp* embryos with anti-dl and anti-histone H3 (to detect nuclei), and hybridized them with an antisense riboprobe against *vnd* (Figure 9A). It is clear from looking at the image that these embryos have expanded dl gradients (compare with Figure 1A). However, it is not obvious whether *vnd* expression domain is shifted as a result (see Figure 9B for a plot of the secondary data of this embryo). However, after fitting the *vnd* peaks to the *vnd* canonical gene expression profile (see Figure 9C for example), we find both the ventral and dorsal borders of *vnd* are shifted dorsally compared to wildtype (see boxplot in Figure 9D). For example, from our calculations, the dorsal border of *vnd* in dl-gfp embryos, $x_{d,G}$, is $0.36 \pm 0.02$, while in wildtype, $x_{d,wt} = 0.32 \pm 0.02$. This difference of 4% DV axis length translates to a shift by roughly 2 nuclei.

### 4.3 Quantifying gene expression profiles in other geometries: saggital sections of Drosophila embryos

To demonstrate that our analysis is generalizable to geometries other than the circular cross-section of the *Drosophila* embryo, we present an example of quantification of gene expression profiles in a saggital section of the Drosophila embryo (Figure 10A).

We chose embryos that express a dominant form of the Toll receptor (Toll10B) present in an anterior-posterior gradient (see Section 2 Materials and Methods; Huang et al., 1997). The endogenous, ventral-to-dorsal nuclear gradient of Dorsal is missing in these embryos, replaced instead by an anterior-posterior nuclear gradient of Dorsal. We hybridized these embryos with riboprobes against *sna*, *vnd*, *sog*, and *ind* (Figure 10A) and quantified their gene expression profiles in sliding window around the embryo periphery (Figure 10B; compare to Figure 3). Aside from generating new canonical profiles for each gene, due to the different spacing as a result of changing the geometry, no alterations to the program is required. Note that both *sna* and *ind* occupy the same color channel; our peak-fitting procedure can differentiate between the two peaks (Figure 10C). This example demonstrates the program's ability to quantify profiles in different geometries and to differentiate between two peaks of gene expression in the same color channel.

# 5. DISCUSSION

In this paper, we have presented a protocol that analyzes primary fluorescent image data to produce relationships between fluorescent intensity and location on the periphery of an optical section within a single embryo. These relationships act as secondary data, which are further fit to empirical models of protein and/or gene expression, resulting in tertiary data: parameters physically meaningful to the researcher, such as boundaries of gene expression. Thus, this approach can distill fluorescent image data down to a handful of parameters, making it possible to compare important features of protein or gene expression across a large number of embryos, and/or across sets of embryos of different genotypes.

The method presented here is generalizable to other systems beyond cross-sections of early *Drosophila* embryos. In some cases, the Matlab files may work on other systems with no need for adjustment. One constraint is the pertinent information must be limited to the periphery of the embryo. However, the embryo need not be a perfect circle. Sagittal sections of *Drosophila* embryos work as well (see Section 4.3). Sea urchin embryos, which have more of a pear-like shape, may also work. The Matlab codes to support our protocol, along with a user's manual, is available in the supplementary material.

There are several avenues for improvement. First, with the exception of the Gaussian-fit of the dl nuclear gradient, our empirical model of gene and protein expression is limited to domains that can vary in width and location, but not shape. For example, if one is interested in empirically modeling a gene that changes shape depending on the age of the embryo, such as for *sog* or *zen*, the code must be adjusted to include stage-specific canonical profiles. In previous work, we took this difficulty into account manually, on a

case-by-case basis, and did not incorporate it into the automated peak-fitting code (Reeves et al., 2012). Another example would be a gene expression pattern changing shape in genetically manipulated embryos. In *sna* mutants, *sog*, *vnd*, and other ventral-laterally expressed genes are derepressed ventrally, drastically changing the shape of their profiles. This problem can be overcome by creating a separate canonical profile for these alternative conditions.

Second, with careful attention to experimental detail and imaging conditions, it is possible to perform semi-quantitative analysis of fluorescent images (see also Section 3.2.1). In semi-quantitative analysis, not only are spatial patterns accurate, but background-subtracted fluorescent intensities are directly proportional to protein or transcript levels (Goentoro et al., 2006; He et al., 2008; Liberman and Stathopoulos, 2009; Luengo Hendriks et al., 2006; Reeves et al., 2012). This typically requires all experimental procedures to be performed in one set, all by the same user, and microscope settings to be constant. Changes in laser power are permissible, as they have a predictable effect on the fluorescent intensity of the image, but must be measured with each imaging session. Additionally, background levels must be measured, through imaging immunostained embryos that lack the protein of interest. The measurements of laser power and background levels must serve as additional inputs to the image analysis code. The automated protocol presented here currently does not support these inputs.

Finally, the procedure for nuclear detection can be improved. At low nuclear densities, such as in earlier timepoints of *Drosophila* development (nc 11 and 12), there is a very high incidence of false detection of nuclei. Manual nuclear identification resolves this issue (Reeves et al., 2012), but is not a part of the code presented here. Furthermore, our

detection of nuclei takes each slice in a z-stack individually. A more comprehensive, three-dimensional nuclear detection algorithm would expand on the accuracy of the code. For example, currently for identification of nascent transcripts, the strength of the nuclear dot present within each slice of a z-stack is plotted and a global trend in the data along the dorsal-ventral axis is inferred based on application of a smoothing function (see Figure 5D). With three-dimensional nuclear detection, a single datapoint for each nascent transcript could be obtained (i.e., two datapoints for transcripts from autosomes; and 1 or 2 datapoints for transcripts from the X, depending on whether the embryo is male or female).

Because our approach also results in a rigorous statistical analysis of the uncertainty in the tertiary data, we can have confidence in our ability to detect subtle phenotypes. We have demonstrated this in two examples using the dl morphogen gradient in the early *Drosophila* embryo as a model system. These subtle phenotypes, not easily detected by eye, are nonetheless very significant, from a statistical standpoint. It is often the case that patterning is robust in developing embryos, making analysis of mutants challenging. Utilizing this method, we can be statistically confident in characterizing small changes in the system.

Another important advantage to the method of using empirical models of protein/gene expression is that we do not neglect the bulk of our secondary data. For instance, one approach to determining gene expression boundaries is to find where the peak of gene expression crosses the half-maximal intensity. In such a case, only two small regions of the secondary data are used: the region at the peak intensity (to determine the maximal value) and the region near the border (to determine the location). Furthermore, in

this scenario, the ability to accurately measure the boundary location becomes quite weak when there is significant noise in the secondary data.

This is clearly not the first research to empirically model protein distributions or gene expression patterns. The Bicoid gradient, which is a protein gradient along the anterior-posterior axis in the early *Drosophila* embryo at the same time as the dl nuclear gradient, has often been empirically modeled by an exponential profile (Gregor et al., 2005; Houchmandzadeh et al., 2002; Manu et al., 2009). However, these choices of empirical models are very specific to the protein being studied. A more general approach, which we employ for gene expression, would be to use a canonical protein profile. Such an approach is mechanism-independent and can be used for any arbitrarily-shaped, but consistent, distribution.

The empirical modeling we present here, when related back to the canonical pattern, results in the definition of physically meaningful parameters, such as shifts in the locations of gene expression boundaries. This is not always the case in similar modeling studies, as another general approach that has been used to model gene expression patterns is fitting to Fourier series (Umulis et al., 2010). For this reason, when the mechanism is unknown, we argue for the utility of our approach based on empirical fitting. However, the future challenge is to use our quantitative measurements (i.e., the physically meaningful parameters) to help build mechanistic models of gene expression and infer biomolecular properties associated with the embryo. Ultimately, the overlying goal is to use such quantitative analysis to understand how gene expression in the embryo is controlled at the level of the gene regulatory network (Ay and Arnosti, 2011; Perkins et al., 2006).

FIGURES AND LEGENDS

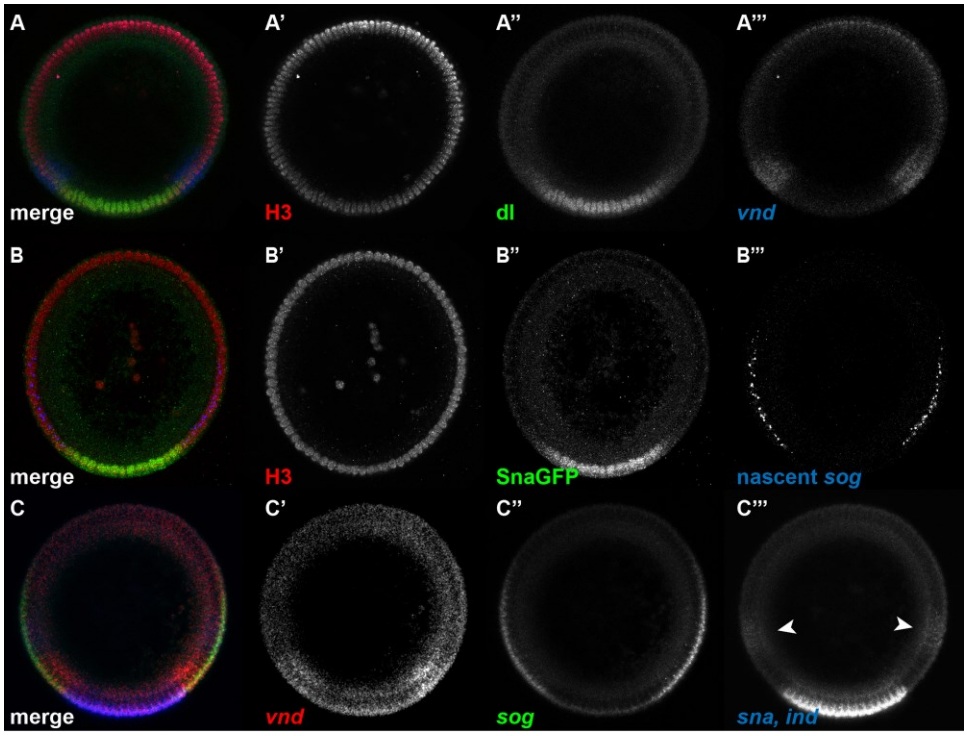



**Figure 1: Examples of *Drosophila* embryo cross-sections analyzed with the methods presented in this paper.**

**(A)** Embryo stained with anti-histone H3 antibody (red; A') and anti-dl antibody (green; A''), and hybridized with a *vnd* antisense RNA probe (blue; A'''). This embryo represents an example of a nuclear stain, a nuclear protein to be detected, and an mRNA expression pattern.

**(B)** Embryo stained with anti-histone H3 antibody (red; B') and anti-GFP antibody (green; B''), and hybridized with a *sog* intronic probe (blue; B'''). This embryo represents an example of a nuclear stain, a nuclear protein to be detected, and a nascent transcript pattern.

**(C)** Embryo hybridized with antisense RNA probes against vnd (red; C'), sog (green, C''), and ind (blue; white arrowheads in C''') and sna together (blue; C'''). This embryo represents an example of four mRNA expression patterns to be detected, including a single color channel that has two mRNA expression patterns simultaneously.

Each of these embryos is a manual cross section of a nc 14 wildtype embryo, with ventral side oriented down. The embryo in (A) is also present in Figures 1, 2, 4, 6, and 7. The embryo in (B) is also present in Figure 5.

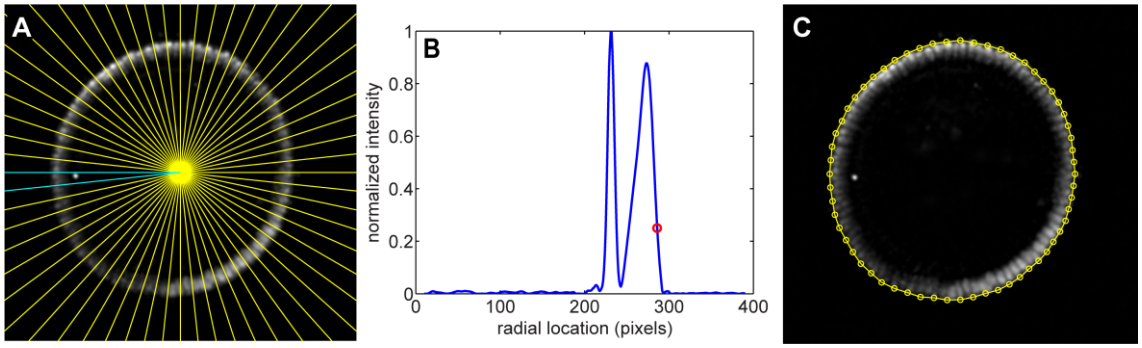



**Figure 2: Finding the periphery.**

**(A)** The embryo image is divided into 60 domains radiating out from the image center (yellow lines). The two cyan lines represent the boundaries of the domain shown in (B).

**(B)** Intensity of the pixels in the domain shown in (A) as a function of radial distance from the center of the image. The red circle denotes the outermost point in which the image intensity rapidly drops from high to low, signifying the periphery of the embryo in this domain.

**(C)** For each domain shown in (A), the points where the boundary of the embryo is determined to be.

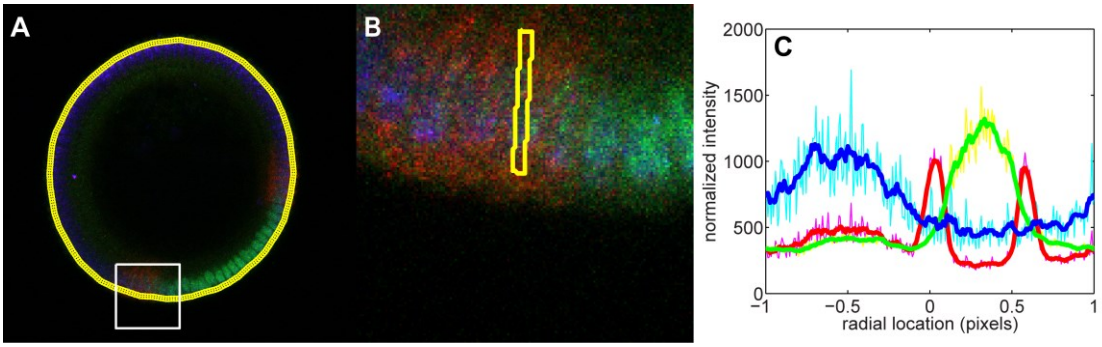



**Figure 3: Measuring average intensity around the periphery of the embryo.**

**(A)** RGB image of an embryo for which the boundary has been calculated. The yellow ring encircling the embryo is the dense mesh of 300 points. The white square denotes the portion of the image shown in (B).

**(B)** Higher magnification of the area shown in (A). The yellow quadrilateral denotes the domain for which a point of average intensity of each color channel is taken. 300 such quadrilaterals around the periphery of the embryo are used to generate the plot in (C).

**(C)** Average intensity in the quadrilaterals as you go around the periphery of the embryo (see (B) for an example). The colors correspond to the color channels in (A), (B). The lighter colors are the raw data, and the darker, thicker curves are the smoothed data. Before midline centering, 0 is taken to be the bottom-most point of the embryo in the image.

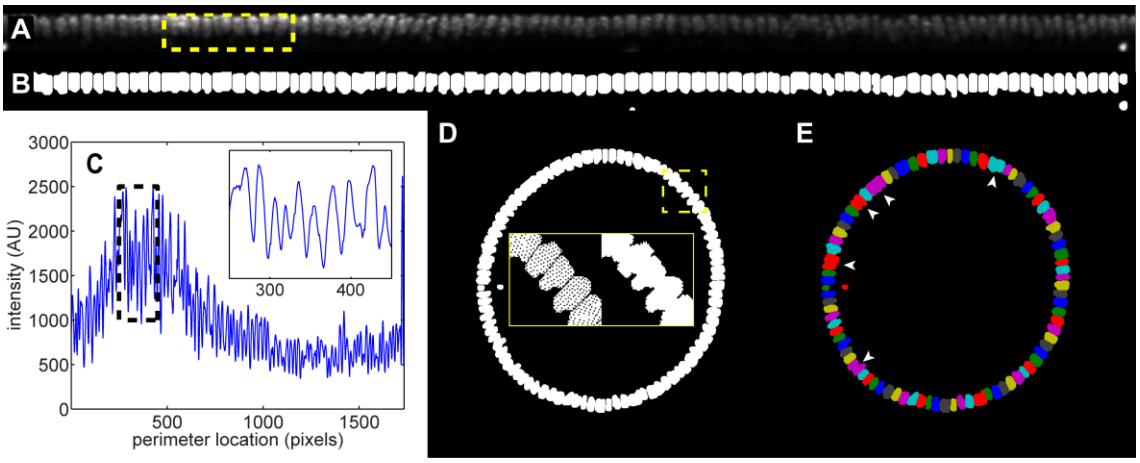



**Figure 4: Locating nuclei.**

**(A)** Unrolled strip of nuclei. The yellow box represents the nuclei in the inset in (C).

**(B)** The nuclear mask of the unrolled strip, with nuclei in white.

**(C)** The image shown in (A), averaged in the vertical direction to reveal a 1D representation of the nuclei. Black dotted box: the area shown in the inset. The inset reveals that the individual nuclei can be resolved in this way.

**(D)** Nuclear mask shown in (B), re-mapped back onto the original 2D image. Yellow dashed box: area depicted in insets. Insets: zoomed-in view of re-mapped nuclei before (left) and after (right) filling in holes in the nuclei.

**(E)** Each nucleus, after being mapped back onto the original image, is identified by a separate color. That way, even if two nuclei overlap in the original image, they can be distinguished using the unrolling technique. Where the primary data provide poor contrast between nuclei, adjacent nuclei may be incorrectly lumped together (arrowheads).

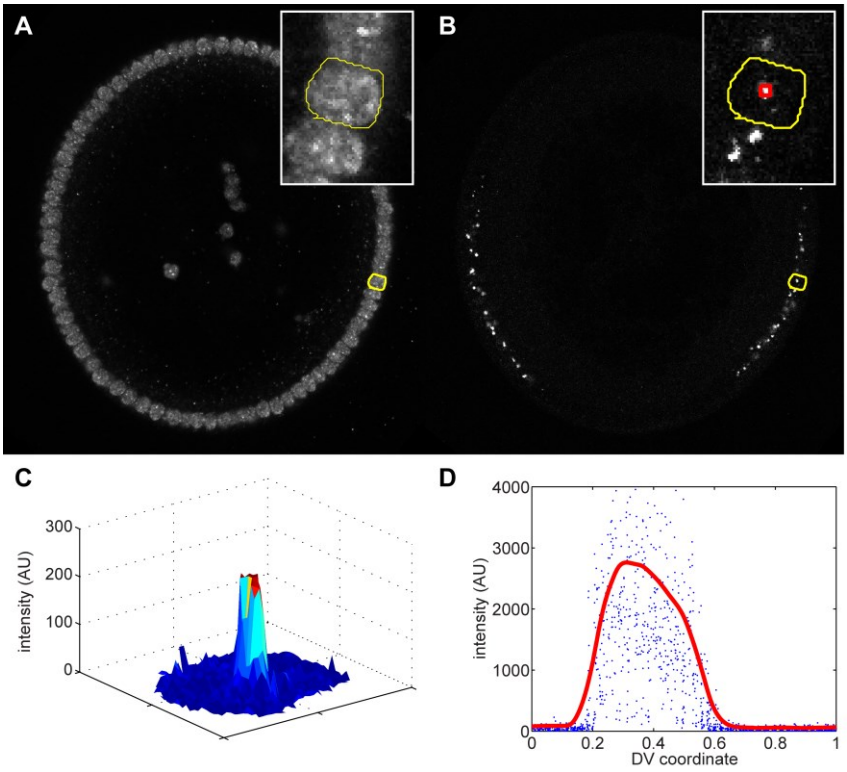



**Figure 5: Measuring nuclear dots.**

The embryo shown in this image has been immunostained with an anti-histone antibody to mark the nuclei and an intronic RNA probe against *sog*, which detects nascent transcripts of *sog* (nuclear dots). Cross-section view with ventral-side down.

**(A)** Nuclear channel. Yellow curve outlines a single nucleus. Inset: magnification of image near the highlighted nucleus (outlined in yellow).

**(B)** Intronic probe channel. The yellow curve outlines the same nucleus as in (A). Inset: magnification of image near the highlighted nucleus (outlined in yellow). Red box: the 5-by-5 neighborhood centered on the max intensity pixel.

**(C)** Surface plot representation of the intensity of the intronic probe channel for only the highlighted nucleus.

**(D)** Measurement of all intronic probe intensities as a function of DV location (blue dots). Data from each slice of a z-stack are plotted together; a single nascent transcript may be represented multiple times if identifiable in multiple slices. Red curve: smoothed version of the blue dots. The smoothed version identifies the trend.

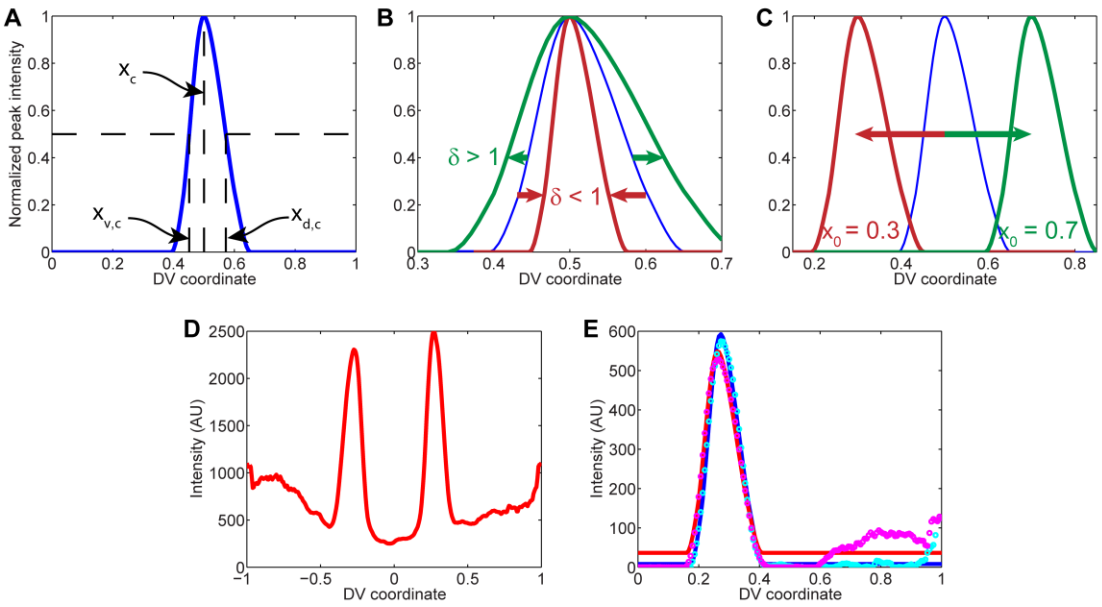



**Figure 6: Peak fitting.**

**(A)** Example of canonical gene expression pattern, using *vnd*. Horizontal dotted lines: half-maximal intensity. $x_c$ is the location of the peak in intensity of the canonical pattern, $x_{v,c}$ is the location of the ventral border of the canonical pattern, and $x_{d,c}$ is the location of the dorsal border of the canonical pattern.

**(B)** Effect of changing the stretching factor, $\delta$. When $\delta < 1$, the pattern shrinks about its peak location (rust-colored curve), and when $\delta > 1$, the pattern expands (green curve). The blue curve is the unstretched canonical pattern for *vnd*.

**(C)** Effect of changing the peak location, $x_0$. When $x_0$ is set to 0.3, the pattern shifts ventrally (rust-colored curve), while increasing $x_0$ causes the pattern to shift dorsally (green curve).

**(D)** Secondary data of *vnd* pattern from embryo in Figure 1A.

**(E)** Fits of the stretched and shifted canonical peak of *vnd* expression to the two peaks in *vnd* from part (D). Cyan and magenta circles are the right and left halves of (D), respectively. The blue and red solid curves are the fits of the canonical pattern to the measured right- and left-side patterns, respectively.

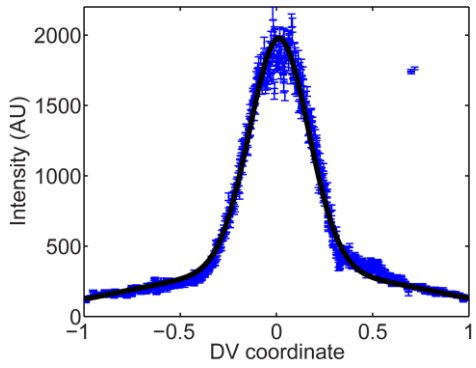



**Figure 7: Gaussian fitting.**

The dl nuclear gradient from the wildtype embryo in Figure 1A is plotted, with errorbars of one standard deviation, as secondary data. The black curve is the best-fit curve according to Equation 1.

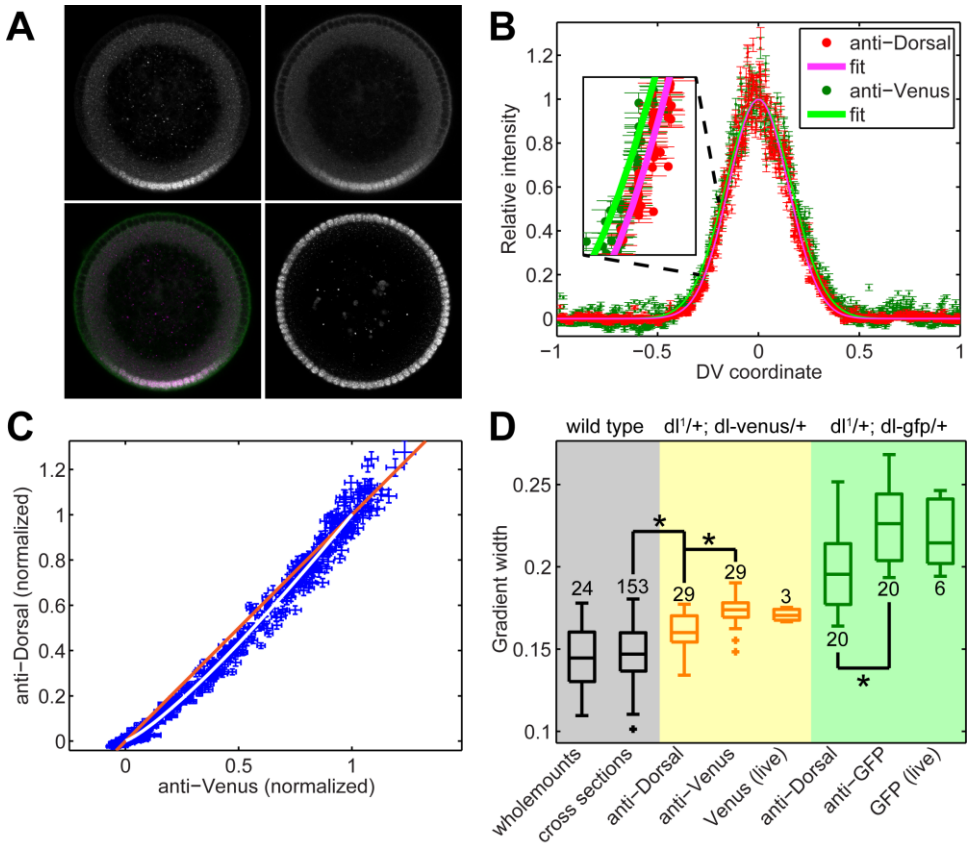



**Figure 8: Expansion of the dl gradient in embryos carrying a copy of dl-Venus.**

**(A)** A *dl-venus* embryo stained for anti-dl (upper left panel), anti-Venus (upper right), and anti-histone H3 (lower right). The lower left panel is a merge between anti-dl and anti-Venus images.

**(B)** Secondary data of the anti-dl (red dots) and anti-Venus (dark green dots) vs. DV axis location. The solid red and solid bright green curves are the fits to the empirical model of the dl gradient (Equation 1). Inset: a higher magnification of the ventral-lateral portion of the two gradients. Errorbars denote one standard deviation.

**(C)** Direct plot of anti-dl vs anti-Venus for each nucleus in the embryo in (A). White curve: plot of the fit of anti-dl vs the fit of anti-Venus. Orange line: 45$^{o}$ line.

**(D)** Boxplot of gradient widths of several different cases. Gray: wildtype embryos. Yellow: *dl-venus* embryos. Green: *dl-gfp* embryos. The latter two genotypes have been measured with anti-dl, with anti-Venus, and live using the native fluorescence of the Venus or GFP tag. Stars indicate statistical significance. Plus signs indicate outliers. Numbers indicate sample size.

Wholemount data from Liberman and Stathopoulos, 2009. All other data from Reeves et al., 2012. 

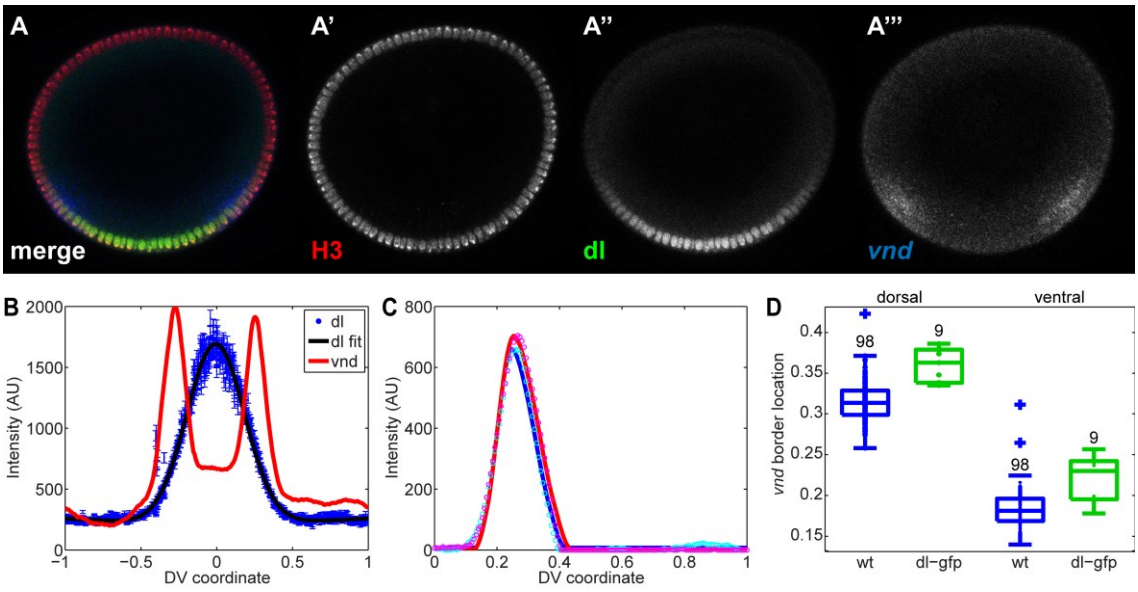



**Figure 9: Expansion of the *vnd* dorsal border in embryos with expanded dl gradients.**

**(A)** *dl-gfp* embryo stained with anti-histone H3 antibody (red; A') and anti-dl antibody (green; A''), and hybridized with a *vnd* antisense RNA probe (blue; A'''). **(B)** Plot of the secondary data from the dl-gfp embryo shown in (A). **(C)** Fits of the *vnd* peaks for this embryo. Cyan and magenta circles are the right and left halves of (B), respectively. The blue and red solid curves are the fits of the canonical pattern to the measured right- and left-side patterns, respectively. **(D)** Boxplot of *vnd* dorsal and ventral border locations in wildtype embryos (blue) and dl-gfp embryos (green). Each dot represents an embryo. Plus signs indicate outliers. Numbers indicate embryo sample size.

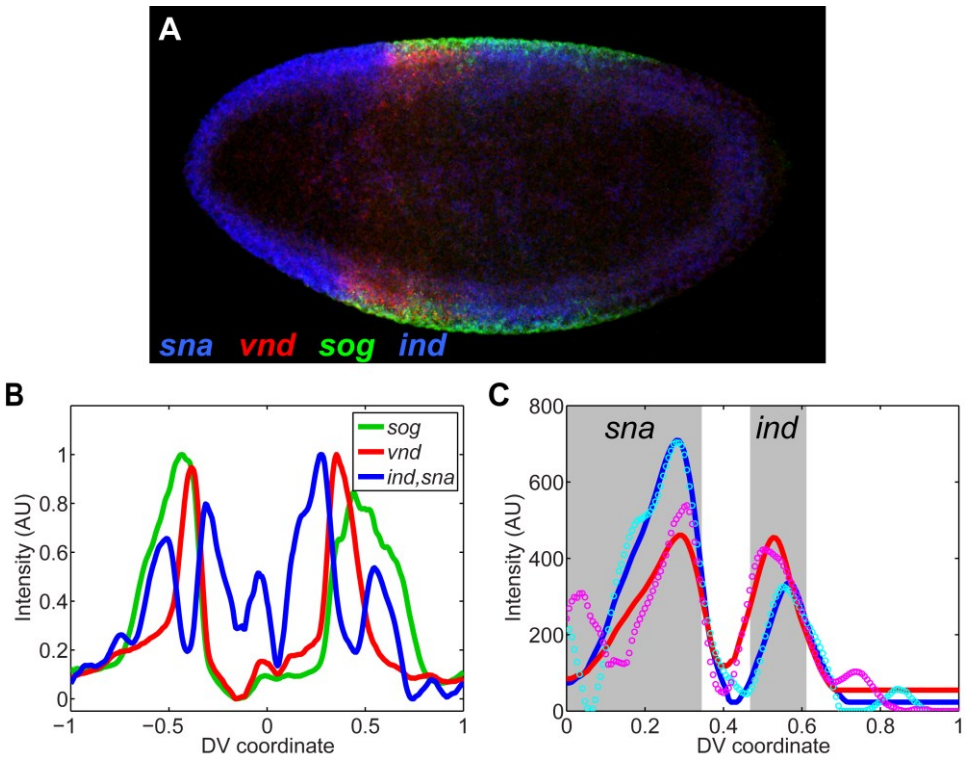



**Figure 10: Quantification of gene expression profiles in a saggital section.**

**(A)** Saggital section of F10 embryo hybridized with probes against sna (blue), vnd (red), sog (green), and ind (blue). Anterior is to the left, and posterior is to the right.

**(B)** Plot of the expression profiles of the genes from embryo in (A). Zero corresponds to the anterior pole (detected manually); one corresponds to the posterior pole.

**(C) F**itting two canonical profiles in the same color channel. The canonical profiles of *sna* and *ind* used for this fit are specifically for F10 embryos. Cyan and magenta circles are the right and left halves of the sna/ind plot in (B), respectively. The blue and red solid curves are the fits of the canonical pattern to the measured right- and left-side patterns, respectively. The gray boxes signify the expression domains of *sna* and *ind*, separated in space.

CHAPTER IV

# The role of FGF signaling in guiding coordinate movement of cell groups: guidance cue and cell adhesion regulator?

This chapter was written with Young-Kyung Bae, Snehalata Kadam, and Angelike Stathopoulos, and published in *Cell Adhesion and Migration,* 6(5):397-403 in 2012.

ABSTRACT

Cell migration influences cell-cell interactions to drive cell differentiation and organogenesis. To support proper development, cell migration must be regulated both temporally and spatially. Mesoderm cell migration in the *Drosophila* embryo serves as an excellent model system to study how cell migration is controlled and influences organogenesis. First, mesoderm spreading transforms the embryo into a multilayered form during gastrulation and, subsequently, cells originating from the caudal visceral mesoderm (CVM) migrate along the entire length of the gut. Here we review our studies, which have focused on the role of fibroblast growth factor (FGF) signaling, and compare and contrast these two different cell migration processes: mesoderm spreading and CVM migration. In both cases, FGF acts as a chemoattractant to guide cells' directional movement but is likely not the only signal that serves this role. Furthermore, FGF likely modulates cell adhesion properties since FGF mutant phenotypes share similarities with those of cell adhesion molecules. Our working hypothesis is that levels of FGF signaling differentially influence cells' response to result in either directional movement or changes in adhesive properties.

# INTRODUCTION

Cell migration is a fundamental developmental process that involves interplay between extracellular signaling molecules, cell surface receptors, and intracellular signal transduction pathways (Ridley et al., 2003). Movement of cells is often directional, with cells sensing the appropriate direction of migration based on recognition of region-specific cues (Rorth, 2011). During embryonic development, cell migration is a very influential process as it results in rearrangement of cells from one part of the embryo to another, effectively controlling cell-cell interactions to drive cell differentiation and organogenesis. *In vitro* studies using cell culture have provided many mechanistic insights into cell migration. However, *in vivo* studies undeniably provide additional insight into the role of the natural environment.

Many studies in a number of model organisms have provided important knowledge regarding how groups of cells move in a coordinate fashion to influence morphogenesis during development (Friedl and Gilmour, 2009; Montell, 2006). In the zebrafish, several signaling pathways including FGF influence collective migration of cells of the lateral line primordium to control both morphogenesis and migration (Aman and Piotrowski, 2011; Friedl and Gilmour, 2009). In the neural crest of vertebrates, it is clear that communication between cells within the migrating collective is necessary for the group of neural crest cells to move; as a result of these cell-cell interactions, contact-dependent cell polarity through N-cadherin is modulated to regulate cell movements (Theveneau et al., 2010). Studies of tracheal cell migration in *Drosophila* have shown that FGF signaling influences the collective movement of this cell group; cells with the highest levels of FGF activity take the lead position (Ghabrial and Krasnow, 2006). We propose

that comparative studies of different systems may provide important insight into general mechanisms that guide collective cell migration.

**Strength of the *Drosophila* system for *in vivo* analyses of cell migration**

The fruit fly *Drosophila melanogaster* is a genetically tractable organism that contains many components of mammalian signaling pathways. *Drosophila* has little genetic redundancy compared to vertebrates, and other strengths of this system include the short generation time (10 days) and relatively quick methods for generating transgenics (4 weeks). Therefore, in *Drosophila*, cellular and genetic approaches can be combined to study biological processes that often provide insights into human dysplasia and disease (Bier and Bodmer, 2004).

For example, *Drosophila* is an excellent system to study how FGF signaling supports development. Only three FGF ligands [Pyramus (Pyr), Thisbe (Ths), and Branchless (Bnl)] and two FGF receptors [Heartless (Htl) and Breathless (Btl)] exist in *Drosophila* (Tulin and Stathopoulos, 2010b). Furthermore, we have shown that only three receptor-ligand complexes are active: Pyr and Ths activate Htl, while Bnl activates Btl (Kadam et al., 2009). In contrast, over 120 FGF-FGFR combinations presumably function in vertebrates (Zhang et al., 2006). In *Drosophila*, the Htl fibroblast growth factor receptor (FGFR) is encoded by a single exon so it is likely that Pyr and Ths activate the same isoform, making this the first pair of invertebrate FGFs to bind the same FGFR isoform (Tulin and Stathopoulos, 2010b). In addition, Pyr and Ths exhibit significant homology to vertebrate FGFs, specifically, to the FGF8 family (Stathopoulos

et al., 2004). Given all this information, the *Drosophila* model system offers a great potential for studying FGF signaling and why ligands often act concurrently.

Here we discuss two FGF-dependent cell migrations, where in both cases Htl FGFR is expressed in the migrating cells, during *Drosophila* embryogenesis. First, FGF signaling through Htl FGFR controls how mesoderm cells come in contact with the ectoderm and promotes mesodermal cell movement as one migrating collective (McMahon et al., 2008; Wilson et al., 2005). Secondly, at a later stage of embryogenesis, Htl-dependent FGF signaling directs a long-distance migration of two cell clusters called caudal visceral mesoderm (CVM), required for proper gut formation (Kadam et al., 2012). These two cell migration events appear quite different: in one case, a tube of cells collapses to a mound of cells, which then spreads into a monolayer such that every cell directly contacts the ectoderm; and in the other case, two distinct groups of cells move coordinately on the left and right sides of the embryonic body from the posterior of the embryo toward the anterior. Nevertheless, FGF signaling supports these two movements in what appears to be a similar manner, supporting both directional movement and also, possibly, modulation of cell adhesion state (Kadam et al., 2012; Kadam et al., 2009; McMahon et al., 2010; McMahon et al., 2008).

We suggest that levels of FGF ligands influence whether FGF signaling acts to regulate chemoattraction (far from the FGF source/low FGF concentration) versus cell adhesion (close to the FGF source/high FGF concentration). As a cell is attracted to move towards the correct 'position', it would make sense that cell adhesion is

upregulated to help the cell remain where it should be. Below we review the relevant data that lead us to propose this model.

# DISCUSSION

## Case I: Mesoderm spreading during *Drosophila* gastrulation

Migration of mesoderm cells during gastrulation is an important step for the regional specification of various mesodermal derivatives (Frasch, 1995). It has been appreciated for a while that FGF signaling is required to support mesoderm cell movement (Beiman et al., 1996; Gisselbrecht et al., 1996; Murray and Saint, 2007; Wilson et al., 2005), but its role in this process was not understood until recently. Htl FGFR is expressed in the migrating mesoderm and two ligands (Pyr and Ths) are expressed in the ectoderm.

To provide insight into the role of FGF in supporting mesoderm spreading during gastrulation, we devised an imaging protocol that allows examination of the movement of hundreds of mesoderm cells deep within *Drosophila* embryos during gastrulation (Supatto et al., 2009). Embryos with ubiquitously expressed histone H2A-GFP were imaged and nuclei of mesoderm cells were tracked, using methodology that we developed (Supatto et al., 2009). Tracking data was transformed into cylindrical coordinates to fit the body plan of the embryo: collapse of the mesodermal tube and intercalation movements occur in the radial direction; dorsal spreading occurred in the angular (azithumal) direction; whereas a strong movement along the length of the embryo was correlated with germband elongation. These studies showed that movement of mesoderm cells during gastrulation is directed and appears highly organized (e.g., the angular position at the end of the migration process is twice that at the start, for each and every cell) (McMahon et al., 2008). Moreover, through live imaging of wildtype

embryos, we identified that cell movements relating to collapse, spreading, and monolayer formation are distinct, as they do not overlap temporally.

The fact that these movements occur in a stepwise manner suggested to us that different molecular signals control each step. Our data showed that FGF signaling through the Htl FGFR controls one of the earlier steps, organized collapse of the mesodermal tube onto the underlying ectoderm (Figure 1A,B) (McMahon et al., 2008). This organized and symmetric collapse is crucial for the subsequent movements. This step positions all mesoderm cells close enough to the ectoderm to support the subsequent spreading, perhaps so that they might also receive additional guidance cues. In FGF mutants, tube collapse often occurs randomly, and this likely contributes to the variability of mesoderm spreading defects observed (Figure 1C). For instance, if the invaginated tube collapses to the right or left, then a more severe 'lumpy' mesoderm phenotype is observed (Figure 1D). However, in a FGF mutant where the tube by chance collapses symmetrically at the midline (as in wildtype), then the mesoderm spreading defect is quite subtle (Figure 1E) (McMahon et al., 2008).

Based on our combined approach of live imaging, cell tracking, and quantitative analyses, we determined that mesoderm cells move as two behaviorally distinct cell populations in *htl* mutant embryos. It was not appreciated before our study that a subset of mesoderm cells maintains their ability to migrate coordinately in the absence of FGF signaling – those cells in contact with the ectoderm exhibit a dorsally directed migration as in wildtype (i.e., movement in the angular direction) (Figure 1C, red cells). In contrast, those cells located at a distance from the ectoderm, which originate from the

upper part of the invaginated tube, appear lost and undergo random movements (Figure 1C, blue cells) (McMahon et al., 2008). However, even in mutants, if cells from the upper part of the invaginated tube happen to come close to the ectoderm, those cells are able to join the migrating collective and also to move in a directed fashion (Figure 1C, cyan cells). Quantitative analysis of cell tracking data was necessary to provide this insight.

In addition, we found that cell intercalation influences spreading (McMahon et al., 2008), but it does so most clearly during cell monolayer formation, the last step of the mesoderm spreading process (McMahon et al., 2010). These final intercalation events simply turn a multi-layered mesoderm organization into a monolayer without any additional dorsally-directed movement. This process is not a convergent extension, but more analogous to 'zippering'.

FGF signaling is required to guide cell movement radially toward the ectoderm (McMahon et al., 2010; McMahon et al., 2008). First, our data suggest that expression of Thisbe, specifically, in ventral regions of the ectoderm controls collapse (McMahon et al., 2010). FGF signaling through Thisbe likely promotes directional movement of mesoderm cells in the tube toward the ectodermal source of this ligand, to ensure symmetrical collapse of the invaginated tube (McMahon et al., 2010). Protrusions have been observed that extend from cells located in the tube toward the ventral ectoderm, using electron microscopy (Wilson et al., 2005), which argues for a chemoattractive mechanism supporting collapse. Secondly, our recent analysis suggested that expression of both ligands, which collectively encompasses the entire ectoderm, influences efficient

monolayer formation (in the radial direction) at the final stage of this mesoderm spreading process. FGF mutants (Figure 1E) and integrin mutants exhibit defects in monolayer formation (McMahon et al., 2010). This shared phenotype supports the view that this last stage in the process requires modulation of mesoderm cell adhesion to the substrate to support monolayer formation. It had been proposed that a straightforward FGF chemoattraction guides dorsally-directed movements in the angular direction, since the localized expression pattern of the FGF ligand Pyramus resides within the dorsal ectoderm (Kadam et al., 2009). However, cells are able to migrate dorsally even in *pyr* mutants (McMahon et al., 2010), indicating that Pyr does not provide cues necessary for dorsally-directed migration.

These new data lead us to propose that FGFs have a distinct function at low versus high concentrations: at low concentration they act in a chemoattractive manner to direct cell movement/orientation required for symmetrical collapse of the invaginated mesoderm tube; whereas at high concentration they act to increase cell adhesiveness to support short-range intercalation movements required for monolayer formation.

**Case II: Caudal visceral mesoderm migration required for gut formation**

Using a similar approach, we also recently investigated the role of Heartless, Pyramus, and Thisbe in supporting migration of another group of cells in the *Drosophila* embryo, CVM cells (Campos-Ortega and Hartenstein, 1997; Kadam et al., 2012). While a role for CVM cell migration in gut formation has been appreciated, little is known about

how these cells accomplish their migration, the longest migration in all of *Drosophila* embryogenesis (Ismat et al., 2010; Kusch and Reuter, 1999; Urbano et al., 2011).

The CVM migration consists of distinct steps. First, the CVM cluster at the posterior end separates into two symmetric groups: left and right. Subsequently, these two groups of ~30 cells each undergo coordinate and directed movement toward the anterior of the embryo (Figure 1F early, 1G). The migration ensues over six hours and throughout the entire course the two separate groups migrate synchronously. This process is necessary to position CVM cells along the entire length of the developing gut (Figure 1F late, 1H). Lastly, at the end of their migration, CVM cells fuse with fusion-competent myoblasts to form the longitudinal muscles that ensheath the gut (Lee et al., 2006).

Our working hypothesis has been that CVM migration, like mesoderm spreading, is a multi-step process as different inputs likely influence cells' movement during the course of their long-distance migration. To start, our studies have focused on the role of FGF signaling in guiding this migration as (i) in FGF mutants the longitudinal visceral muscle fibers, which arise from CVM cells, are absent (Kadam et al., 2012; Mandal et al., 2004), and (ii) FGF signaling components are expressed in the CVM and the trunk visceral mesoderm (TVM). Htl is expressed in the migrating CVM cells (Mandal et al., 2004), and its ligands Pyramus and Thisbe are expressed within TVM, a substratum ('track') upon which CVM cells migrate (Kadam et al., 2012; Stathopoulos et al., 2004). The TVM is present as two bands on either side of the embryo, with each band serving as a track for the migration of one cluster of CVM cells.

To obtain insight into the role of FGF signaling during CVM migration, in a recent study (Kadam et al., 2012), we investigated whether FGF guides directional movement of CVM cells, as these cells stay closely associated with the TVM (Figure 1F). Our results suggest that FGF signaling functions in a chemoattractive manner to guide CVM cell migration and also supports cells' survival. In the absence of FGF signaling, cells from the right and left sides of the embryo veer off course, detach from their respective TVM, and converge at the midline. In some cases, cells cross over completely to the alternate side, which is a phenotype not observed in wildtype (Figure 1I early,1J). In addition, overexpression of Pyr and/or Ths FGFs at an ectopic location, at the ventral midline, redirects CVM cells toward this source of ligand. Furthermore, most CVM cells eventually die in FGF mutants (Kadam et al., 2012; Mandal et al., 2004). While this might relate to some checkpoint mechanism that ensures that cells that have gone off-track are eliminated, our data support the view that FGF signaling also likely supports cell survival (Figure 1I late, 1K). Ectopic expression of ligands at a distance can rescue cell viability even if migration remains 'off-track' (Kadam et al., 2012).

However, even in the absence of FGF signaling, CVM cells still initiate anteriorly-directed forward movement, albeit somewhat misdirected and slow (Figure 1I late,1K). While CVM cells in the FGF mutants are disorganized, perhaps through lack of adhesive properties, they ultimately move forward as long as they are kept alive. Therefore, FGF-independent signals likely exist that also guide anterior movement.

We propose that FGF signaling supports several roles throughout the six hours that CVM cells undergo their long-range migration. Initially, wildtype FGF signaling

acts in a chemoattractive manner to recruit CVM cells onto the TVM tracks, upon which cells migrate. In the FGF mutants, cells veer off–course and cross over at the midline; this never happens in wildtype embryos (Kadam et al., 2012). However, expression of ligands is found along the length of the TVM, so it is not clear how FGF ligands would support chemotactic movement toward the anterior in the absence of a gradient. Processing or some other modification of ligands may support graded FGF activity to support forward movement of cells once at the TVM. Alternatively, it is possible that FGF signaling, in this context with CVM cells at the TVM, acts as a 'permissive' signal to allow cells to effectively sense other signals that may influence anteriorly-directed movement. For example, once CVM cells reach the TVM the role of FGF may be to simply keep CVM cells 'on track', possibly through regulation of cell adhesion properties, so that they remain in range to receive other guidance cues (Reim et al., 2012).

**How do cell collectives migrate in a coordinate fashion?**

Whereas one signal may suffice to guide migration of small groups of cells, more complex mechanisms likely safeguard proper migration of larger groups of cells. In addition to signals influencing direction of migration, it is probable that cells within each group must coordinate with each other to ensure that each migrating collective moves in a directed fashion.

Coordination between cells in a migrating collective may require physical association between them, either stable or transient, and/or chemical signaling. During

the mesoderm spreading process, cells are closely associated and likely linked by adherens junctions as well as gap junctions (Tepass and Hartenstein, 1994). However, these structures remain to be defined in terms of their components (which involve a number of different proteins), their prevalence and dynamics, and their role in supporting cell movement. Nevertheless, these junctions certainly have important roles during morphogenesis (Harris et al., 2009). Cell-cell interactions also occur between neighboring CVM cells, but cells within the migrating collective appear loosely associated. As CVM cells also interact closely with the TVM, we hypothesize that CVM-TVM cell-cell interactions play a significant role in supporting CVM cells' anteriorly-directed movement. More careful analysis of the physical associations of homotypic (mesoderm-mesoderm) as well as heterotypic (mesoderm-ectoderm or CVM-TVM) cell-cell interactions should provide important insights. In addition, synchronous migration of the two CVM clusters is abolished in FGF mutants, suggesting a possible novel role for FGF in long-range cell-cell communication. Potential influences to be investigated include regulation of cell adhesion properties, direction of movement, orientation/number of cell projections, cell division, and/or cell viability. The complexity of collective migration is highlighted here as each of these features involves multiple proteins and layers of regulation.

**Distinct and overlapping functions of FGF ligands**

While a very impressive analysis of all vertebrate FGF-FGFR interactions was recently completed in which the binding specificities of ligand-receptor interactions were

examined in tissue culture (Zhang et al., 2006), how this relates to *in vivo* processes, for the most part, is undetermined. Our studies have focused on obtaining this exact information, to define specific roles for each *Drosophila* FGF *in vivo* (Kadam et al., 2012; Kadam et al., 2009; McMahon et al., 2010; McMahon et al., 2008). We have demonstrated that Pyramus and Thisbe ligands have both overlapping as well as distinct functions within the *Drosophila* early embryo: (i) Thisbe controls collapse of the invaginated mesoderm during gastrulation; (ii) both ligands are required to form a proper mesoderm monolayer as the end result of mesoderm spreading during gastrulation; (iii) subsequently, primarily Pyramus alone is required for differentiation of dorsal mesoderm lineages; and (iv) lastly, both ligands work together to support migration of CVM cells later in embryogenesis.

In our most recent study of CVM cell migration (Kadam et al., 2012), we found that ectopic expression of Pyramus and Thisbe together (at the ventral midline in embryos lacking endogenous ligand expression) caused a severe migration defect: CVM cells were essentially recruited to the ectopic site and then stalled. This result brings up the interesting possibility that the combined activity of both ligands is distinct from having either one, because expression of each ligand individually did not support this effect. This led us to propose that FGF ligand heterodimers can support a distinct function possibly through differences in binding affinity, stability, and/or recruitment of cofactors. FGF ligand homodimers bound to FGFR were crystallized and the structure obtained suggested that heterdimeric binding is also possible (Plotnikov et al., 2000).

Why particular developmental processes depend on Pyramus and/or Thisbe is not understood. These molecules may activate distinct intracellular signaling downstream of Htl-activation to support different cell behaviors, for example cell migration versus cell differentiation (Franzdottir et al., 2009). However, it does not appear that Pyramus and Thisbe have dedicated functions. For instance, Pyramus supports differentiation of dorsal somatic mesoderm lineages in the embryo by supporting cell differentiation (i.e., transcriptional response) (Kadam et al., 2009; Klingseisen et al., 2009), while supporting cell migration for glia associated with neuronal development of the eye at later stages (Franzdottir et al., 2009). Alternatively, FGFs may exhibit different range of action or be subject to different regulation. Regarding this last point, we have determined that these ligands are differentially cleaved and that the C-terminus of Thisbe may function to inhibit activity (Tulin and Stathopoulos, 2010a). *Drosophila*, with a total of three FGF ligands compared to 22+ genes in vertebrates, is an attractive model system to investigate the individual activities of FGFs.

**Could FGF signaling supports cell movement by regulating cells' adhesive state?**

Cell adhesivity may influence cell-cell interactions to help cells move as a single migrating collective, affecting homotypic interactions and/or the ability of cells to interact with the substratum upon which they migrate.

Our analysis, tracking nuclei, examined the mesoderm spreading process following collapse, and suggested mutant cells were more loosely associated with each

other. Results showed that movements of cells that originate from the upper part of the tube, and thus do not contact the ectoderm, were misdirected (appearing random) and encompassed far larger distances than normal (McMahon et al., 2008). Furthermore, when Pyramus or Thisbe ligands are ectopically expressed in the mesoderm (essentially increasing FGF signaling), tracking analyses have found that all mesoderm cell movement is halted (Amy McMahon and A.S., unpub. obs.), perhaps through increased adhesion. Our data, following collapse, is consistent with the view that lack of FGF activity results in weak mesoderm-mesoderm cell-cell associations, possibly 'rescued' by contact with the ectoderm, whereas too much FGF signaling supports cell-cell associations that are too strong and actually hinder motility (Figure 2). Another study has shown that at an earlier stage in the absence of FGF signaling, EMT is delayed due to defects in E-Cadherin redistribution (Clark et al., 2011), which is a molecule that can influence cell adhesion properties. An interesting future direction would be to investigate whether mesoderm and ectoderm cells' adhesion state changes during the various steps of this mesoderm migration process (i.e., EMT, collapse, spreading, and monolayer formation).

Along these lines, when both ligands are expressed in combination (but not individually) within the CVM, cell movement is halted. For the few cells that are able to 'break free' from the collective, they appear to migrate just fine. One interpretation of this result is that ectopic expression of ligands results in cessation of movement as CVM cells become too 'adherent' to each other. It will be of great interest to examine how the

cell adhesion properties of these stalled cells are altered by overexpression of both ligands.

In addition, in subsequent studies of the mesoderm spreading process, we have found that FGF mutant phenotypes share similarity with those of genes that influence cell adhesion – the Rap1 GTPase and the βPS integrin Myospheroid (McMahon et al., 2010). While the Rap1 GTPase influences both collapse and monolayer formation (similar to FGF), the integrin Myospheroid is specifically required for the final step of this process, monolayer formation. It is possible that proper monolayer formation requires a substantial increase in cell-cell adhesion between mesoderm and ectoderm.

Our working hypothesis is that FGF signaling serves multiple roles to support cell movement. At lower levels, FGF ligands may serve as chemoattractants but once levels are raised, for instance when migrating cells approach the ligand source, then a secondary function of FGF signaling acts to increase cell-cell adhesion properties (Figure 2). While a role for FGF signaling in modulating cell adhesion to support cell movement remains unclear in the field, experiments using the *Drosophila* model system have the potential to provide necessary insight. The relative ease of genetic manipulation and live imaging of *Drosophila* shows promise for the study of the complex and dynamic processes that relate to collective cell migration.

FIGURES AND LEGENDS

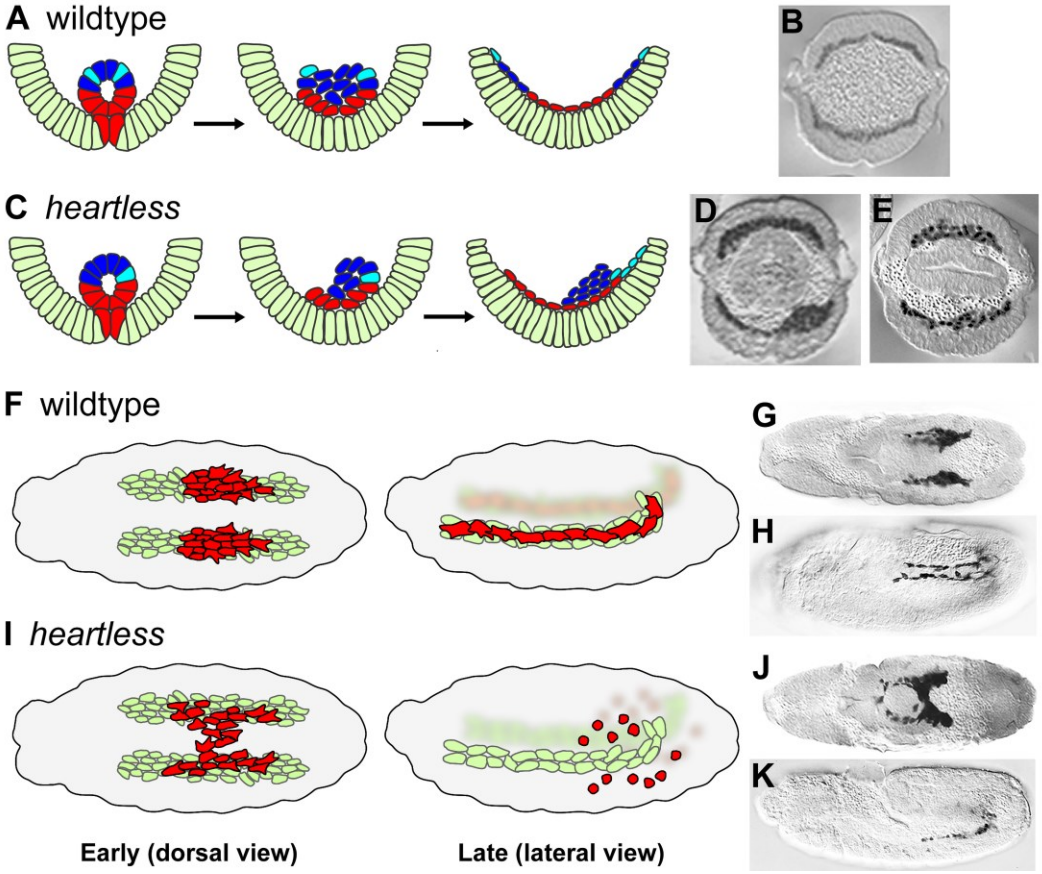



**Figure 1. Comparison of cell movements in wildtype and *heartless* mutant embryos.** (A,C) Schematic based on published results (McMahon et al., 2008); (B,D-E) cross-sections of anti-Twist staining of wildtype and *htl* mutant embryos. (A,B) In wildtype, all mesoderm (red/blue/cyan) cells contact the ectoderm (light green) and are able to spread dorsally to form a monolayer. (C) In *htl* mutants, only the subset of cells (red/cyan) that contact the ectoderm undergoes directed movements. Depending on how the tube collapses, the mutant phenotype can be severe (D) or subtle (E). (F,I) Schematic based on published results (Kadam et al., 2012); CVM reporter *croc-lacZ* in wildtype

(G-H) and *htl* mutant (J-K) embryos stained with anti-βgal oriented with anterior to the left. (F,G) The dorsal view of wildtype at stage 11 shows the two distinct, symmetrical clusters of CVM cells (red) migrating on the two bands of TVM cells (light green). (F,H) At stage 13, the lateral view reveals complete CVM migration with cells evenly distributed along the TVM. (I,J) In *htl* mutant embryos, CVM cells are intermixed in early migration. (I,K) Later stages of mutant embryos illustrate CVM cell death and loss of contact with the TVM. Panels (B, C, and D) and (G and J) were reprinted with permission from McMahon et al., *Science* 2008 and Kadam et al., *Development* 2012, respectively.

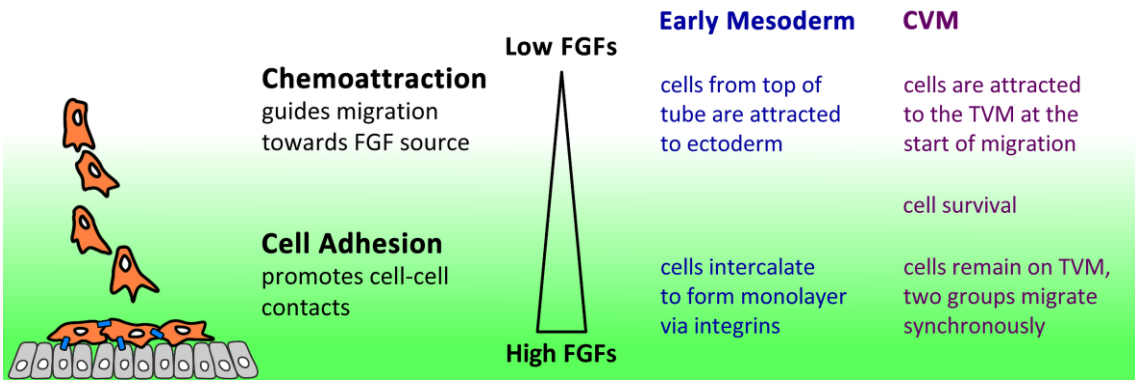



**Figure 2. Model of FGF's dual function.** Data from recent studies suggest that FGFs are able to function differentially in a concentration dependent manner. The concentration of FGFs is governed by the proximity of responding cells (orange) to the source (gray cells producing FGFs). At a distance, where levels are low, FGFs works as chemoattractant such that cells become polarized and migrate directionally. Once cells are closer to the source, the higher levels of FGFs promote cell adhesion (blue lines).

CHAPTER V

# Screen identifies spatiotemporal expression of proteoglycans Trol and Syndecan is important to support mesoderm development in the *Drosophila* embryo

ABSTRACT

Gastrulation of an early embryo involves coordinate cell movements supported by multiple signaling pathways, adhesion molecules, and extracellular matrix components. Fibroblast growth factors (FGFs) have a major role in *Drosophila* mesoderm migration, however other inputs and/or co-factors to FGF acting in this process are unclear. We performed an ectopic expression screen for genes that impact mesoderm cell migration and identified eleven genes that disrupt migration including the FGF ligand Pyramus, α-integrins, E-cadherin, Cueball, EGFR and JAK/STAT signaling components, as well as several enzymes. Here we investigated one gene isolated encoding the heparan sulfate proteoglycan (HSPG) Terribly reduced optic lobes (Trol), ortholog of mammalian HSPG Perlecan. Collectively, our data support the view that HSPGs modulate ligand range of action at multiple FGF-dependent processes: we found that Trol is able to function at a distance to reinforce long-range ligand-receptor interactions such as those controlling large cell movements during early mesoderm collapse and caudal visceral mesoderm (CVM) migration. Conversely, Sdc but not Trol is required in cases where short-range signaling is acting, such as in defined domains of pericardial cell differentiation or to support cell intercalations in forming the mesoderm monolayer. We propose these HSPGs, one extracellular and the other membrane-bound, differentially regulate FGF signaling by localizing ligands in different ways.

# INTRODUCTION

Embryonic development requires integration of multiple complex processes such as cell movement, proliferation, and differentiation. Important regulators of these processes are signaling pathways. To ensure proper execution of these processes, it is important that signaling pathway activation is tightly regulated (Solnica-Krezel and Sepich, 2012). It is therefore important to understand how individual signaling pathways are regulated as well as the integration of multiple inputs.

The *Drosophila* fibroblast growth factors (FGFs) Pyramus (Pyr) and Thisbe (Ths) and their receptor Heartless (Htl) have been previously shown to function in supporting mesoderm migration during gastrulation (Bae et al., 2012; Winklbauer and Muller, 2011). FGF signaling regulates the mesoderm's radial movement through Rap1 GTPase and integrins (McMahon et al., 2010; McMahon et al., 2008). This includes the initial collapse of the furrow and the intercalation of cells to form the monolayer. In *fgf* mutant embryos, cells at the dorsal most part of the furrow cannot properly collapse to reach the ectoderm following epithelial-mesenchymal transition (EMT) or intercalate at the end of the process to support formation of a monolayer. However, cells in contact with the ectoderm are able to migrate laterally even in the absence of FGF signaling. Therefore, other inputs besides FGF are also likely important for guiding directional movement of mesoderm cells during gastrulation.

Specifically, we hypothesized that additional signaling pathways and/or regulators of cell adhesion may act to support mesoderm migration at gastrulation. To investigate how cells were still able to migrate in the absence of FGF signaling and also to discover

additional components in the FGF pathway, we conducted a screen of genes secreted or membrane-tethered proteins using an ectopic expression approach to identify those that impact mesoderm migration at gastrulation. We postulated that important signals guiding this process normally would be differentially expressed in tissues in the embryo, either in the mesoderm or ectoderm, in order to provide positional information to guide mesoderm cell movements. The UAS/GAL4 system was used to ectopically express candidate genes in either the presumptive mesoderm or ectodermal tissues. In this way, a cell-surface and secreted (CSS) insertion collection, first used in a neuronal pathfinding screen (Kurusu et al., 2008), was used to identify eleven genes that impact *Drosophila* gastrulation when ectopically expressed. We focused analysis on one gene isolated in this screen encoding a heparan sulfate proteoglycan (HSPG), Terribly reduced optic lobes (Trol), due to previous research linking HSPGs and FGFs. Crystal structures have revealed that HSPGs bind to the FGF ligand and receptor as a heterotrimeric complex (i.e., FGF-HSPG-FGFR) (Pellegrini et al., 2000). It has been proposed that HSPGs facilitate ligand-receptor interaction and/or stabilize the FGF-FGFR dimer complex.

HSPGs are comprised of a core protein attached with highly modified heparan sulfate glycosaminoglycan (HS GAG) side chains that are known to regulate multiple signaling pathways during development (Lin, 2004). There are only four known core proteins in *Drosophila*: transmembrane Syndecan (Sdc), two membrane-anchored glypicans Dally and Dally-like (Dlp), and the extracellular matrix protein Trol. Trol is the homolog of mammalian Perlecan (Pcan), and several lines of evidence support the view that Pcan promotes multiple pathways including FGF signaling in vertebrates (Farach-Carson et al., 2014). For instance, *in vitro* experiments measured a gradient of FGF-2 and

correlated its levels with Pcan, and pERK, a signal measuring activation of the Ras intracellular signaling pathway downstream of FGFR activation (Wu et al., 2014). Studies in the developing mouse heart show specific Pcan modifications are required for binding of different FGF-FGFR complexes (Allen and Rapraeger, 2003). In *Drosophila*, studies of the larval lymph gland have suggested that the Trol HSPG sequesters FGF signaling in this tissue (Dragojlovic-Munther and Martinez-Agosto, 2013). However, *trol* mutant phenotypes in the early *Drosophila* embryo have not previously been investigated.

# RESULTS

*Ectopic expression screen identifies genes in multiple pathways affecting mesoderm development*

Presumptive mesoderm cells are initially specified prior to gastrulation in ventral regions of the embryo (Solnica-Krezel and Sepich, 2012). These ventral cells undergo shape changes that cause a furrow to form, comprising the presumptive mesodermal domain. Apical constriction of cells drives their invagination during which time a tube is formed on the inside of the embryo. Cells within this invaginated tube undergo EMT driving their collapse upon the inner surface of ectodermal cells. These presumptive mesoderm cells migrate first in the dorsal direction and subsequently in the radial direction to establish a single layer of mesoderm cells on the inside of the embryo (Figure 1A).

In order to elucidate potential signaling pathways and adhesion molecules that influence mesoderm migration, we conducted a screen of a library comprised of 311 insertions at the presumed 5' end of genes encoding cell surface or secreted (CSS) factors (Figure 1B). These lines were previously selected to help with identification of extracellular-acting signaling molecules and used in a screen of neuronal targeting (Kurusu et al., 2008). Using these fly stocks in the current study, we aimed to identify novel regulators of mesoderm spreading during gastrulation. To this end, genes were overexpressed using Gal4 drivers that support expression in the mesoderm (Twi-Gal4) or ectoderm substratum (69B-Gal4). Twenty-four insertions were identified that caused lethality upon ectopic expression in the mesoderm and/or ectoderm (see Table 1).

Next, we screened these candidates in order to determine if lethality could have been caused by defects in mesoderm migration. Genes were selected that are expressed in the embryo at the appropriate stage: stages 5-10 encompassing invagination through complete mesoderm spreading. If expression was observed at this stage within the migrating mesoderm and/or ectoderm substratum, analysis was prioritized. Concurrently, we examined embryo cross-sections for spreading defects in mutant backgrounds. Single null mutants were examined if available and, if not, deficiency chromosomes deleting the gene in question (along with many others) were examined. We reasoned that genes normally acting to support the mesoderm spreading processes would exhibit mutant phenotypes. Screening in this manner identified eleven genes of interest.

*Signaling components, adhesion molecules, and modifying enzymes identified as regulators of mesoderm migration during gastrulation*

FGF ligand, integrins, and cadherin genes were identified by the screen, which were expected and support previous roles for these signals in supporting mesoderm migration during gastrulation (McMahon et al., 2010; Oda et al., 1998). An insertion upstream of the FGF ligand Pyr (GS22603) resulted in embryonic lethality upon ectopic expression with the 69B-Gal4 driver (data not shown). A previous study has characterized the role of the FGF ligand Pyr during mesoderm migration (Kadam et al., 2009). In addition, prior studies have also identified a role for the β-integrin Myospheroid (Mys) in this process, demonstrating that it is required solely for monolayer formation at the end of the process following spreading of cells on the ectoderm (McMahon et al., 2010). In the current screen, two alpha integrins, α-PS3 (Scab; EP2591) and α-PS5 (GS12413), were identified,

which may act with Mys. Integrins function in tetramers with the binding of two α- and two β-integrins (Bulgakova et al., 2012). Initial analysis of α-PS3 and α-PS5 shows expression in the mesoderm and multilayer spreading defects in embryos mutant for the integrin (Figure 2A-D). *Drosophila* contains three additional alpha-integrins that are present during early mesoderm development (Figures S1A-C). It is possible that these α-integrins all support mesoderm migration during gastrulation (see Discussion). Lastly, E-cadherin (Ecad, Shotgun) was isolated. Cadherins are considered necessary for their function in adhesion and EMT (Clark et al., 2011). Ecad is expressed in the ectoderm at gastrulation when mesoderm migration occurs (Oda et al., 1998), and mutants show a non-monolayer mesoderm (Figure 2E,F).

Because these genes had already been implicated in controlling cell movements during gastrulation, we focused on analysis of other genes that might provide novel insights into this process. We identified an insertion (EY1263) near the *cueball* (*cue*) gene, which encodes a membrane-bound protein that is EGF-like and contains LDLR repeats. It is expressed in the mesoderm and embryos lacking *cue* exhibit a multilayer phenotype (Figure 2G,H).

Only two genes induced embryonic lethality when overexpressed in both the mesoderm and ectoderm (Figure S1D,E,M,N). Both of these genes are secreted factors and ligands influencing signaling pathways: Unpaired (Upd; G17133) regulates the JAK/STAT pathway, whereas Vein (Vn; GS12044) regulates EGFR signaling. While previous studies have focused on *upd* during heart diversification (Johnson et al., 2011), a role in the early mesoderm at gastrulation has not been identified. Upd is expressed in ectodermal stripes and mutant embryos results in a multilayer phenotype (Figure 1I,J).

Modification to JAK/STAT signaling only had mild effects on mesoderm migration (Figure S1F-L).

The second secreted molecule that resulted in embryonic lethality when overexpressed in both the mesoderm or ectoderm was Vein, an epidermal growth factor receptor (EGFR) ligand. Vn is expressed in the ectoderm and *vn* mutants have a strong multilayer spreading phenotype (Figure 2K,L). Another EGF pathway component, Argos (Aos), was also identified in the screen. Aos is expressed in the ectoderm (Figure 2M) and the deficiency covering *aos* presented a less severe, but uneven, mesoderm spreading phenotype (Figure 2N). Since two components of the EGFR pathway were identified in the screen, we also examined the phenotype associated with the receptor itself (Shilo, 2005). EGFR is upregulated in the mesoderm when spreading is complete, and expressing its dominant negative form in the mesoderm resulted in a strong multilayer phenotype (Figure 2O,P). However, *egfr* mutants and ectopic expression of the EGFR dominant negative in the ectoderm had little to no effect on the mesoderm even though EGFR is present in the ectoderm at earlier stages (Figure S2O-T).

It is possible that the JAK/STAT and EGFR signaling pathways are active in the mesoderm during migration. Future studies may distinguish direct from indirect roles; for instance, these pathways may regulate gene expression and/or protein distributions of other genes within the ectoderm.

Our screen also identified two enzymes of unknown function: predicted sulfotransferase *CG9550* (GS18034) and galactosyltransferase *CG34056* (GS11028). Analyses of these genes show mesoderm expression and spreading defects when analyzed in the context of deficiency chromosomes (Figure 2Q-T). However, more than twenty

genes were uncovered by these large deletions, therefore it is unclear whether these phenotypes relate to the genes in question (Figure S1U-Y). Nevertheless, these enzymes could potentially function in the synthesis and/or modification of proteoglycans, which were also found in the screen (see below).

*Comparing proteoglycans in the Drosophila embryo*

The *Drosophila* genome contains four HSPGs (Lin, 2004) and two predicted chondroitin sulfate proteoglycans (CSPGs) (Song et al., 2012). In our screen, we identified two proteoglycans, one of each type: Trol (HSPG) and PTP99A (CSPG), which led us to more closely interrogate these genes as well as others of this class for a role in mesoderm migration. Sdc, an HSPG, has previously been identified to play a role in mesoderm development in the embryo (Knox et al., 2011). We examined the remaining HSPGs Dally and Dlp and putative CSPG Kon-tiki (Kon) for their embryonic expression (Figure 3). All genes, except *kon*, are maternally deposited and are present during mesoderm migration. In addition, Trol, Dally, and Kon are expressed in what appears to be the caudal visceral mesoderm (CVM), another group of FGF-dependent migrating cells (Kadam et al., 2012).

The *trol* locus spans ~75 kB and includes as many as 58 exons encoding 22 unique polypeptides (Figure 4A). Overexpressing Trol or *trol* RNAi constructs in either the ectoderm or mesoderm results in moderate spreading defects (Figure 4B-E). However, germline clones devoid of both maternal and zygotic (m-z-) *trol* transcripts exhibit mesoderm tube collapse defects that result in a severe multilayer mesoderm phenotype (Figure 4F). The maternal contribution is sufficient to rescue the collapse and non-

monolayer (Figure 4G). Importantly, this phenotype is similar to that found in embryos lacking FGF signaling (McMahon et al., 2008). We therefore investigated whether FGFR receptor activation was possible in the absence of Trol, by assaying for dpERK in cells within the mesoderm having migrated to the dorsal-most position (Figure 4H); dpERK is a measure of FGFR RTK intracellular signaling activation (Gabay et al., 1997b). dpERK is absent in embryos from *trol* germline clones (Figure 4I). Furthermore, when Trol is overexpressed in the ectoderm or mesoderm, dpERK is expanded or ectopically expressed, respectively (Figure 4J,K). Embryos lacking *trol* also had an overall reduction of dpERK in ectodermal cells (Figure 4I), indicating Trol may also support other signaling pathways (e.g., EGFR; Gabay et al., 1997a). Furthermore, *trol* germline clones exhibit a "tail-up" phenotype suggesting abnormal TGF-β signaling (compare Figures 4L,M) (Ferguson and Anderson, 1992). These data support the view that Trol modulates FGF signaling, and possibly other pathways as well, during early mesoderm development.

*Trol and Sdc have different roles in embryonic development*

Because both Trol and Sdc (Knox et al., 2011) exhibit phenotypes that affect the mesoderm of early embryos, we interrogated their expression patterns during early mesoderm development to provide more specific insights into their functions. Both genes are expressed ubiquitously at low levels, likely because they are maternally-expressed, except at two stages when localized expression was observed. Once the furrow is formed, *trol* is upregulated in the ventral-most cells where the mesoderm will collapse onto the ectoderm (Figure 5A). Conversely, *sdc* becomes localized to the ectoderm later when the

mesoderm intercalates to form a single layer of cells (Figure 5B). Its ectodermal expression domain is similar to that of the FGF ligand Ths (Kadam et al., 2009).

In accordance with the *sdc* expression pattern, removing Sdc from the ectoderm or overexpressing Sdc in the mesoderm results in severe mesoderm spreading defects. In contrast, the inverse perturbations (i.e., *sdc* RNAi in the mesoderm and Sdc overexpression in the ectoderm) result in only mild spreading defects (compare Figure 5C with D, Figure 5E with F). However, overexpressing Sdc in either domain causes an overall decrease in dpERK staining (Figure 5G,H). This is similar to *trol* germline clones (loss-of-function phenotype) and suggests lack of dpERK staining upon Sdc overexpression relates to the ability of Sdc to potentially compete with Trol. Ectopic expression of other HSPGs Dally and Dally-like display mild to no mesoderm spreading defects (Figure S2A-H). While ectopic expression of CSPG Ptp99a resulted in a multilayer, removing *ptp99a* in the embryo had little effect (Figure S2I-K). Kon was not examined since this gene is not expressed until later embryonic stages, and therefore does not regulate mesoderm migration.

Pericardial and dorsal somatic muscle cells derived from the dorsal mesoderm are known to express Even-skipped (Eve) (Frasch, 1999), and require proper migration of the mesoderm at an earlier stage. Once mesoderm cells migrate to dorsal regions of the ectoderm, there are induced signals coming from the ectoderm to support Eve+ expression in ten clusters of three cells each spanning the trunk of the embryo (Figure 6A). These differentiation cues include FGF, Wg, and Dpp – all of which have the ability to cooperate with HSPGs (Lin, 2004). We examined if Trol is also required for Eve expression and found that *trol* germline clones had no noticeable defect (Figure 6B). Previous studies,

however, have shown that *sdc* mutants, in contrast, do exhibit subtle defects in Eve expression within dorsal somatic muscle lineages (Knox et al., 2011). Also, overexpression of Sdc in the ectoderm produces strong effects on Eve+ cell number, with multiple clusters containing 5+ cells rather than the normal 3-4 (Figure 6F). Sdc overexpression in the mesoderm only occasionally generated more Eve+ cells, with clusters containing 4-5 cells (Figure 6C-E). Trol ectopic expression, on the other hand, had no effect. These results suggest that Sdc is more important for supporting differentiation of dorsal somatic mesoderm lineages than is Trol.

Of the proteogylcans examined by expression analysis, we found Trol and Kon to be present in a migrating population of cells originating from the caudal visceral mesoderm (CVM). Both *trol* germline clones and *trol* RNAi in the CVM cells resulted in a migration defect in which cells from each of the two migrating collectives merge at the midline (compare Figure 6G with H,I), similar to the phenotype caused by removing FGF signaling (Kadam et al., 2012). These *trol* mutants, along with *kon* RNAi (Figure 6K), also had increased apoptosis as indicated by the punctate spots at the posterior of the embryo. Whether this is due to a role for Trol in supporting cell survival and/or mis-migration is unclear as either could account for the phenotype. Lastly, introducing *sdc* RNAi into CVM cells had no effect (Figure 6J), further supporting the view that Trol and Kon but not Sdc are specifically required in the migrating CVM.

# DISCUSSION

Previously, FGF signaling and the β-integrin Mys were shown to be important for this process as were a few intracellular effectors (Bae et al., 2012; Winklbauer and Muller, 2011), but no other extracellular molecules had been identified. In our screen of cell surface and secreted proteins, in addition to Trol, several α-integrin genes were isolated, some or all of which may bind to known player α-integrin Mys to form tetramers during mesoderm development. While other studies have suggested E-cadherin is necessary for EMT at the onset of mesoderm migration, further studies are needed to determine if cadherins are required for mesoderm spreading itself or if the mesoderm migration defects observed are a consequence of improper EMT. In addition, identification of Cue through the screen is suggestive of the importance of trafficking of signaling components and/or adhesion molecules towards regulation of mesoderm development (Hirst and Carmichael, 2011). The signaling pathways JAK/STAT and EGFR may also function in parallel with FGF to guide the spreading process. Additionally, two enzymes of unknown function were uncovered, CG9550 and CG34056. Other enzymes previously identified to act in mesoderm migration were characterized to function in the biosynthesis of heparan sulfate (HS) side chains found on HSPGs (Lin et al., 1999). This led us to focus on proteoglycans for functional analyses.

Our screen isolated the HSPG Trol and putative CSPG Ptp99a (discussed below). The only other HSPG reported previously to work with FGF during mesoderm development is Sdc (Knox et al., 2011). Comparing extracellular Trol with transmembrane Sdc revealed spatiotemporal differences in expression (Figures 3 and 5, A and B) and non-

overlapping phenotypes in multiple FGF-dependent processes (Figures 4-6). Based on these data, we propose the model that extracellular matrix (ECM)-component Trol is able to promote FGF-FGFR interactions at a distance. This may be through an ability of Trol to support diffusion of the ligand and deliver FGF to the receptor (Figure 7-1). Another hypothesis is that Trol stabilizes FGF and allows presentation of the ligand to be taken up by cells expressing the receptor through cell protrusions such as cytonemes (Figure 7-2) (Roy et al., 2014). On the other hand, membrane-bound Sdc is not able to freely move between cells. It can only function to support short-range FGF signaling in adjacent cells as in cell intercalation (Figure 7-3) and cell differentiation (Figure 7-4).

*Trol requirement in multiple pathways in Drosophila*

While we highlight the role of Trol and Sdc in FGF signaling, our data also suggest that these HSPGs can modulate EGF signaling as indicated by the decrease of dpERK in the ectoderm (Figures 4I and 5G,H) (Gabay et al., 1997a) and TGF-β signaling based on cuticle phenotype (Ferguson and Anderson, 1992) in the case of Trol (Figure 4M). Several studies have linked Trol with FGF signaling as well as other signaling pathways. One of the earlier reports in *Drosophila* demonstrated that Trol is required for FGF signaling through the FGFR Breathless (Btl) and FGF Branchless (Bnl) to support neuroblast proliferation (Park et al., 2003). They also showed that mammalian Perlecan co-immunoprecipitated with FGF-2 and that this interaction can be outcompeted upon addition of heparin. In addition, *trol* mutants displayed higher levels of Hh (Hedgehog) near the morphogen source, suggesting that Trol is required for diffusion of Hh (Park et al., 2003). Another study yielded similar results in the neuromuscular junction by examining Wg-GFP

(Wingless-GFP) of the Wnt pathway (Kamimura et al., 2013). Total Wg levels were not affected in *trol* mutants, however Wg appeared to remain near the presynaptic membranes where it is secreted while the postsynaptic bouton acquired defects analogous to inhibition of Wnt signaling. These reports support the view that a general function for Trol is to effect ligand distribution.

*HSPGs in ECM architecture*

It is possible that HSPGs affect ligand distribution through changing the organization of the ECM, and this can result in positive or negative inputs into signaling pathways (Kim et al., 2011). For example, S2R+ cell culture studies of the HSPG Dlp revealed that it can both enhance and inhibit Wnt signaling, depending on the context (Baeg et al., 2004). Recently, genetic interactions suggest that Trol sequesters the Ths ligand and prevents FGF-dependent differentiation in the larval lymph gland, thus providing an inhibitory role towards FGF signaling (Dragojlovic-Munther and Martinez-Agosto, 2013). However, secreted HSPGs, such as Trol, are also components of the basement membrane and can modify organization of the ECM to create and/or remove barriers during cell migrations (Sarrazin et al., 2011). These lymph glands lacking *trol*, likewise, had defects in the surrounding basement membrane, which were shown to effect Hh distribution (Grigorian et al., 2013). Additionally, the ECM receptor Dystroglycan (Dg) has been shown to bind Trol and is found between the mesoderm-ectoderm interface (Schneider and Baumgartner, 2008). It is expressed in the mesoderm, though *dg* zygotic mutants exhibit normal mesoderm spreading perhaps due to maternal transcript sufficiency

(data not shown). Therefore, *trol* mutants in theory could also indirectly contribute to altered signaling activities at gastrulation due to changes in the ECM structure.

*HSPGs acquire specificity through side chain modifications*

Many other proteins are involved in extending/modifying proteoglycan side chains required for function and specificity (Lin, 2004). Previous studies report that Sugarless (Sgl), Sulfateless (Sfl), and Heparan sulfate 6-O-sulfotransferase (Hs6st), modifiers of HS GAG chains, are enriched in the mesoderm. Sgl and Sfl genetically interact with Htl during mesoderm migration, while Hs6st is essential for FGF signaling in tracheal development (Kamimura et al., 2001; Lin et al., 1999). Our screen identified two enzymes: CG9550, a chondroitin-6-sulfotransferase, and CG34056, a beta-galactosyl-transferase. It is possible that these enzymes modify HSPGs and/or CSPGs in *Drosophila* to influence FGF signaling. In vertebrates, it has been reported that Pcan has both HS and chondroitin sulfate (CS) chains, and that removal of CS chains is required for FGF-2 binding to the HS chains (Smith et al., 2007). It has been previously hypothesized that differences in HS chains are tissue-specific and will determine the potential of proteins to modulate FGF activity (Ornitz, 2000). Therefore, perhaps these enzymes act on Trol and/or Sdc to influence their FGFR activation potentials.

*Extracellular vs. membrane-tethered HSPGs*

In addition to Sdc function in late mesoderm specification (this study; Knox et al., 2011), several other reports implicate membrane-bound HSPGs' involvement in short-range signaling. Axon guidance by Slit/Robo signaling in *Drosophila* embryos requires

two HSPGs, Dlp and Sdc. The distribution and concentration of Dlp and Sdc are discrete in order to generate a distinct spatial field able to direct axonal growth (Smart et al., 2011). Another HSPG, Dally, is necessary in conjunction with BMP signaling for germline stem cell maintenance in *Drosophila* ovaries (Guo and Wang, 2009). This requirement of Dally is limited to the germline only and not the nearby somatic cells, revealing its short range of action. In the vertebrate system, membrane-tethered HS chains are required for FGF signaling in adjacent cells during mouse embryogenesis (Shimokawa et al., 2011). All of these reports emphasize the importance of membrane-bound HSPGs in regulating ligand distribution and limiting signaling activity within a short distance. The property of Trol to be secreted is unique among *Drosophila* HSPGs and provides an extension into the ligand's range of action.

**Conclusions**

We studied the roles of HSPGs Trol and Sdc in FGF signaling during various stages of *Drosophila* embryonic development. Our data demonstrates that secreted Trol is involved in processes where long-range FGF-FGFR interactions are presumably acting, namely collapse of the mesoderm furrow and migration of CVM. Meanwhile, transmembrane Sdc functions when FGF signaling is active within neighboring cells such as in the case of mesoderm cell intercalation and pericardial cell differentiation. Our screen also identified multiple genes in various pathways that affect mesoderm spreading, enforcing the view that many processes are integrated during development. Future studies will reveal how each gene contributes to mesoderm migration.

# MATERIALS AND METHODS

## *Fly strains*

*P{GAL4-twi.G}, w*$^{1}$ (BDSC #914) and *w; P{GawB}69B* (#1774) lines were used in experiments analyzing mesoderm spreading. For screening, females from "virginator" *y*$^{1}$ *w/Dp(2;Y)G, P{hs-hid}Y* (#8846) versions of these *Gal4* stocks were crossed with males from the UAS insertion collection. Wildtype refers to *yw* or *Gal4* lines. Mutant strains were crossed with balancers containing a *lacZ* marker. For germline clones, *trol*$^{G011}$*,FRT.19A/FM7* were crossed with *P{ovoD1-18}P4.1, P{hsFLP}12, y*$^{1}$ *w*$^{1118}$ *sn*$^{3}$ *P{neoFRT}19A/C(1)DX, y*$^{1}$ *w*$^{1}$ *f*$^{1}$ (#23880) and allowed to lay for approximately 12 hours at 25°C. A 2-hour heat shock at 37°C was performed on days 2, 3, and 4. Non-Bar females were then crossed with *y*$^{1}$ *arm*$^{4}$ *w/FM7c, P{ftz/lacC}YH1* males (#616) and collected embryos were analyzed. The *5053-Gal4* driver *w; P{GawB}tey5053A/TM6B, Tb+* (#2702) was used for ectopic expression in the CVM cells and crossed with *bHLH-gap-Venus* (Y-K. Bae and A.S., in review) to detect CVM cells with a GFP antibody. Additional stocks are listed in Supplementary Table 1.

Genes affected by the UAS insertions were confirmed. Sim-Gal4 (S. Crews) or ZenKr-Gal4 (M. Frasch), which support ectopic expression at the embryonic midline or trunk region, were used to drive expression from the insertions and in situ hybridization experiments confirmed ectopic expression of genes (data not shown).

## *In situ hybridization and immunohistochemistry*

Embryos were collected and staged at 25°C or 18°C, and standard protocols for

fixing and staining were used. Antisense RNA probes were made to detect *in vivo* gene expression. Primary antibodies used were rat anti-Twist (1:200; housemade, Pocono), rabbit anti-Eve (1:100; M. Frasch), mouse anti-dpERK (1:150; Sigma), rabbit anti-β-galactosidase (1:200; Molecular Probes), mouse anti-αPS2 (1:10; Developmental Studies Hybridoma Bank), and rabbit anti-GFP (1:2000; Life Technologies).

*Sample preparations and imaging*

For cross-sections, embryos were embedded in araldite (Electron Microscopy Sciences). 10 or 15 μm slices were sectioned using the LKB Bromma 2218 Historange Microtome and mounted in 1:1 araldite:acetone solution. For cuticle preparations, 24-hour old embryos were dechorionated in bleach, devitillinized in 1:1 MeOH:heptane, and mounted in lactic acid. Slides were incubated at 55°C overnight. All images were collected using a Zeiss Axioplan microscope.

FIGURES AND LEGENDS

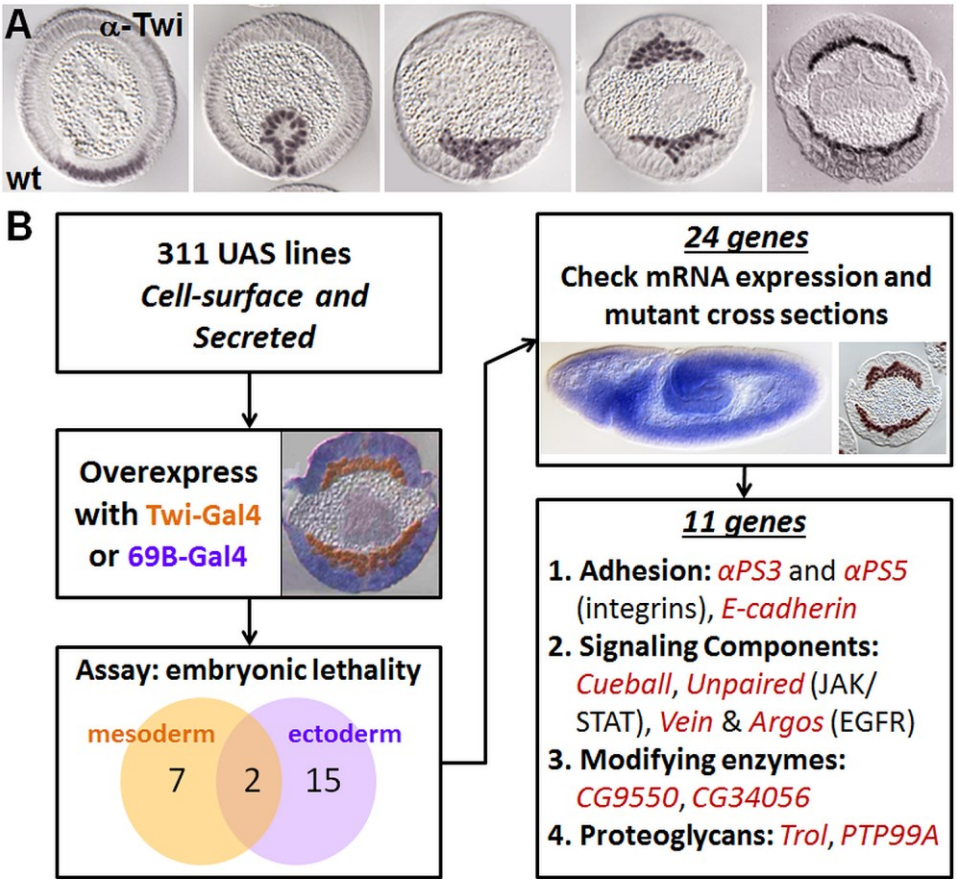



**Figure 1. Eleven genes identified through ectopic expression screen confer mesoderm migration defects.** (A) Cross-section of *Drosophila* embryos stages 5-10 stained with Twist antibody to mark mesoderm cells during development. Ventral side is down. The mesoderm invaginates to form the ventral furrow, which subsequently collapses onto the ectoderm. Lateral migration begins and is completed after intercalation to form a monolayer. (B) Workflow of ectopic expression screen. Cell-surface and secreted (CSS) proteins were overexpressed in the mesoderm using Twi-Gal4 and in the ectoderm using 69B-Gal4. Candidates were narrowed to eleven genes in four different classes by their RNA expression and mutant phenotypes.

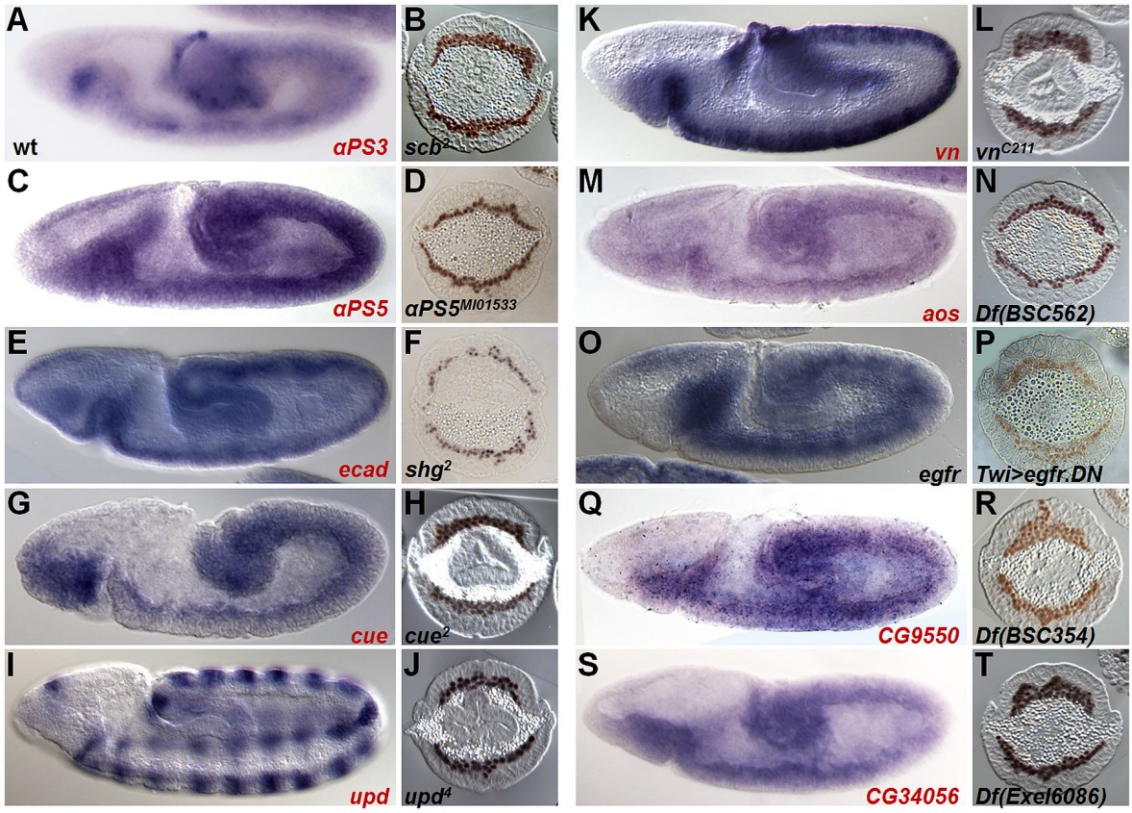



**Figure 2. Expressions and mutant phenotypes of adhesion molecules, signaling ligands, and modifying enzymes isolated from screen.**

RNA expression of stage 8-9 embryos and cross-section of stage 10 mutant embryos for (A, B) αPS3, (C, D) αPS5, (E, F) E-cadherin, (G, H) Cueball, (I, J) Unpaired, (K, L) Vein, (M, N) Argos, (O, P) EGFR, (Q, R) CG9550, and (S, T) CG34056. Mutant embryos are zygotic unless otherwise noted. (P) The dominant negative (DN) form of *egfr* was overexpressed using the Twi-Gal4 driver. In situ hybridization was performed using riboprobes against the indicated genes. Lateral views of whole mount embryos are positioned with anterior facing left and dorsal side facing up. Cross-sectioned embryos stained with α-Twist to mark mesoderm cells. Genes in red denote those isolated from this screen.

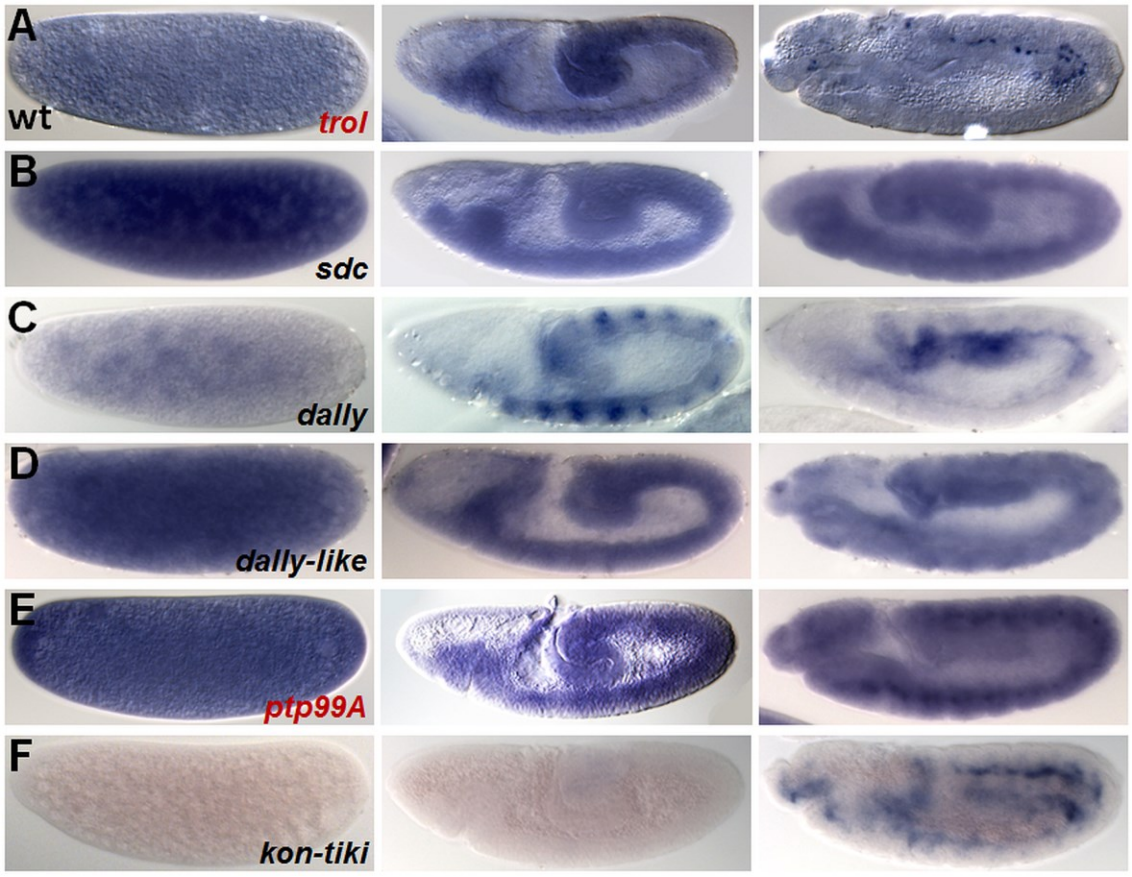



**Figure 3. Embryonic RNA expressions of *Drosophila* proteoglycans are dynamic.** Expression patterns for HSPGs (A) *trol*, (B) *sdc*, (C) *dally*, (D) *dally-like*, and CSPGs (E) *ptp99a* and (F) *kon*. Embryo are staged pre-cellularization (left column), stage 8 (middle), and stage 11 (right).

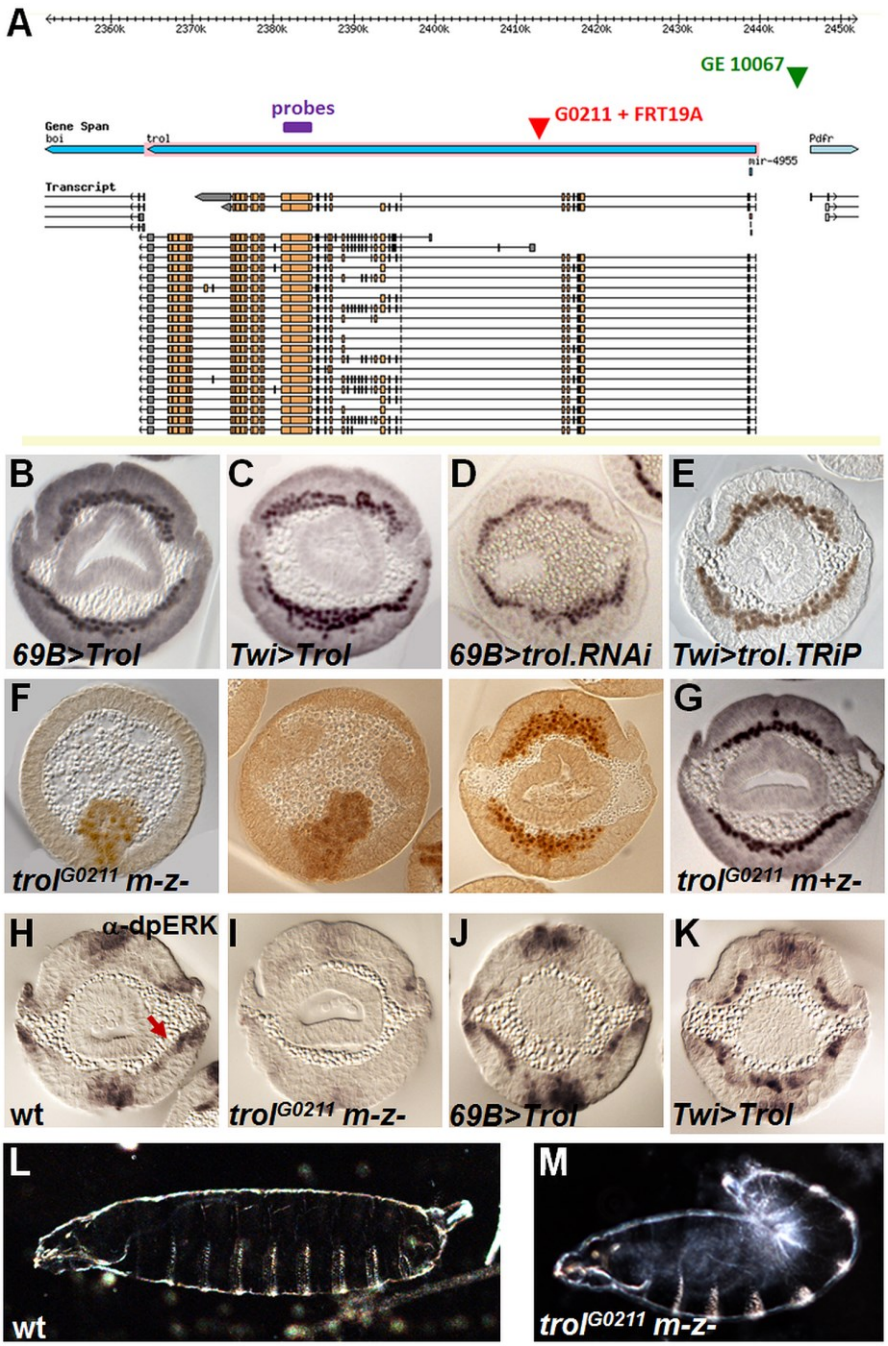



**Figure 4.** ***trol*** **germline clones exhibit defects in mesoderm migration similar to the loss of FGF signaling phenotype.**

(A) Image of *trol* locus obtained from Flybase GBrowse depicting location of the reagents used in this study: GE10067 is a UAS insertion and G0211 is a lacZ insertion that was

recombined with FRT 19A in order to support generation of germline clones. (B) Ectopically expressing Trol in the ectoderm results in embryonic lethality and defects in mesoderm spreading. This multilayer phenotype is also present in embryos overexpressing Trol in the mesoderm (C). Removing Trol by RNAi in the ectoderm (D) or mesoderm (E) yields in mild spreading defects. (F) Only *trol* germline clones exhibit severe defects in mesoderm development, including the ability of the furrow to collapse. (G) Zygotic mutants have normal spreading. (H) dpERK is present at the lateral most mesoderm cells in wildtype embryos. However, dpERK is absent in *trol* germline clones (I) and is expanded when Trol is overexpressed in the ectoderm (J) and mesoderm (K). Cuticles from *trol* germline clones display a "tail-up" phenotype (compare J with I).

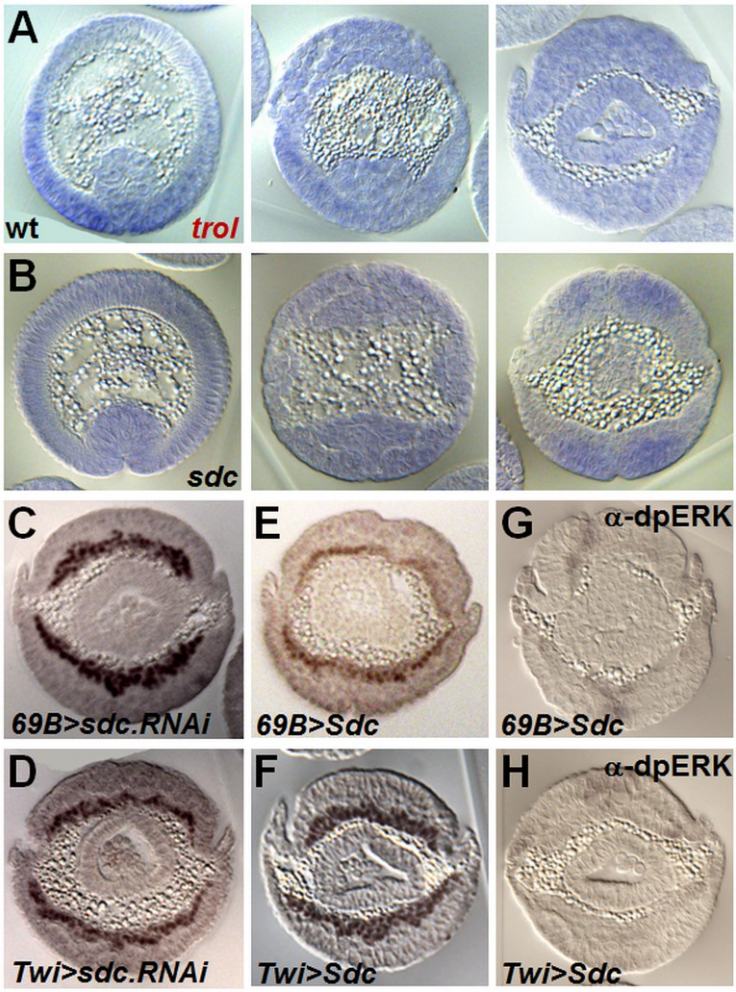



**Figure 5. Syndecan mutant embryos show defects in mesoderm migration.**

Cross-section of embryos stained for (A) *trol* and (B) *sdc* RNA. Note that *trol* is upregulated in the ectoderm cells (ventral-most cells) surrounding the invaginated furrow before collapse. In contrast, *sdc* is localized in the same position but at a later stage when mesoderm cells intercalate to form a monolayer. Overexpressing *sdc* RNAi in the ectoderm (C) gives a slightly more severe phenotype than in the mesoderm (D). Conversely, increasing Sdc in the ectoderm (E) has normal spreading while overexpressing Sdc in the mesoderm results in a multilayer (F). Embryos overexpressing Sdc in the ectoderm (G) and mesoderm (H) both show an overall decrease of dpERK.

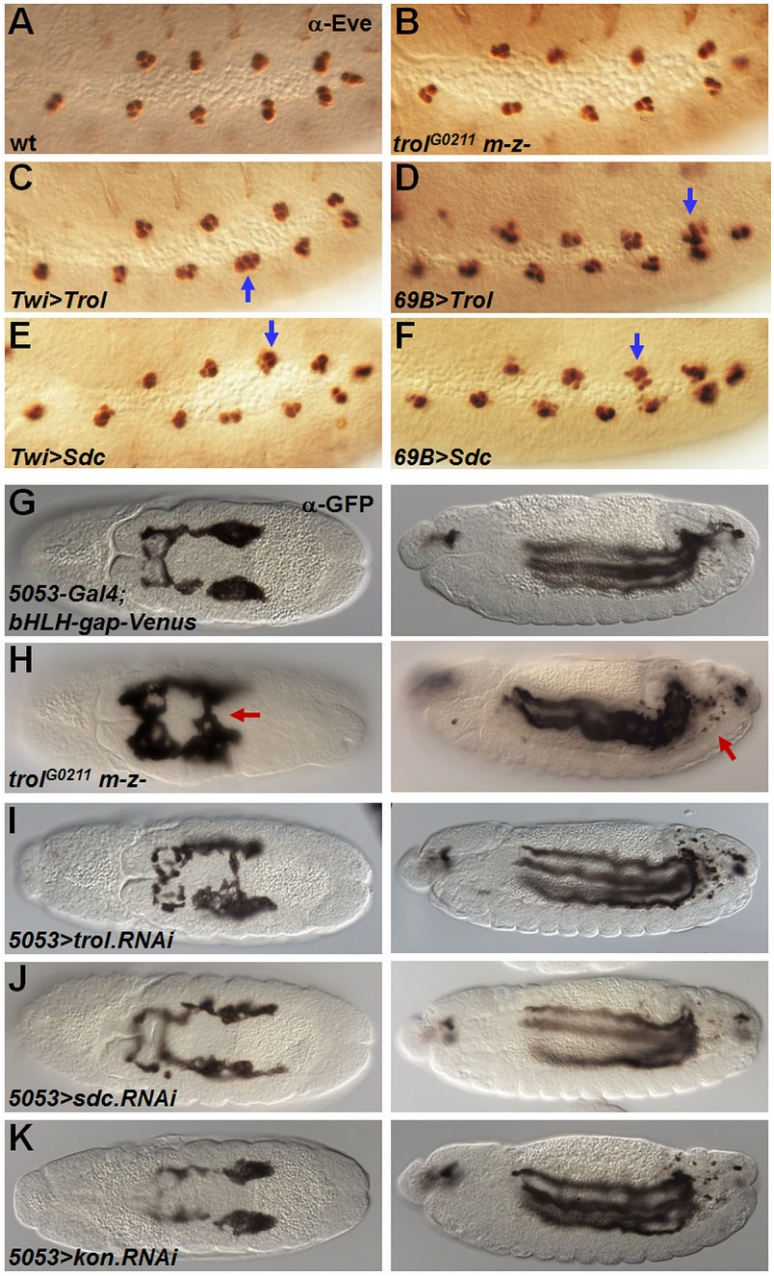



**Figure 6. Sdc affects Eve specification while Trol is required in CVM migration.**

Pericardial cell differentiation marked by α-Eve staining in (A) wildtype, (B) *trol* germline clones, (C) *Twi>Trol*, (D) *69B>Trol*, (E) *Twi>Sdc*, and (F) *69B>Sdc* embryos. Blue arrows show ectopic Eve+ cells. (G-K) Embryos with CVM driver 5053-Gal4 and CVM

marker bHLH-gap-Venus. α-GFP staining marks the CVM in (G) control embryos and (H) *trol* germline clones, and in embryos expressing RNAi against (I) *trol*, (J) *sdc*, and (K) *kon* in CVM cells. Left column is dorsal view of stage 10 embryos with red arrow pointing to merging phenotype. Right column is lateral view of stage 13 embryos with red arrow indicating ectopic cell death in posterior regions.

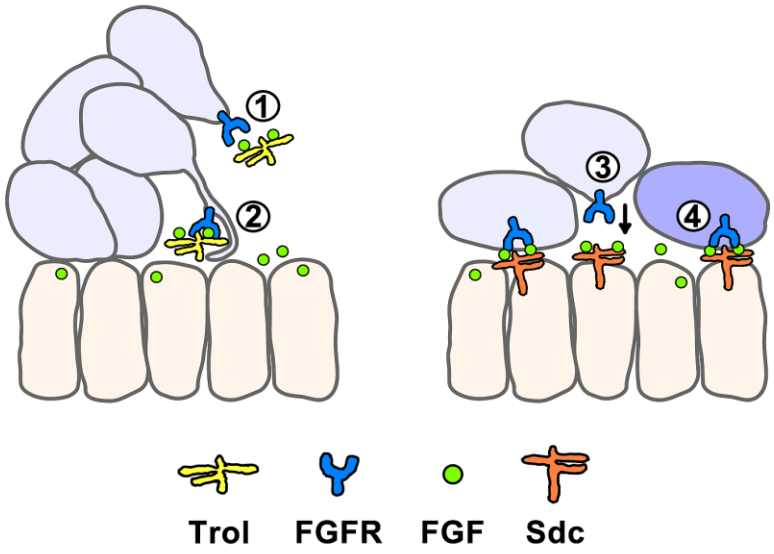



**Figure 7. Trol and Sdc heparan sulfate proteoglycans function with FGF ligands to facilitate long-range versus short-range signaling.**

Trol is secreted into the ECM and has the potential to signal to non-neighboring cells. This may occur through (1) diffusion to target cells and/or (2) the ability to be taken up by target cells via cytonemes. Sdc is bound at the membrane and can thus only signal to adjacent cells to support (3) small movements such as intercalation and (4) cell differentiation.

| Gene ID | Name | | UAS | Lethality |
|---|---|---|---|---|
| CG8084 | Anachronism | Ana | GS 9498 | Twi-Gal4 |
| CG4531 | Argos | Aos | GS 12984 | 69B-Gal4 |
| CG12086 | Cueball | Cue | EY 1263 | 69B-Gal4 |
| CG15013 | Dusky-like | Dyl | GS 20894 | 69B-Gal4 |
| CG3722 | E-cadherin/ Shotgun | Shg | XP d01606 | 69B-Gal4 |
| CG32356 | Ecdysone-inducible gene E1 | ImpE1 | GS 11510 | 69B-Gal4 |
| CG1106 | Gelsolin | Gel | GS 10156 | Twi-Gal4 |
| CG32464 | l(3)82Fd/ Mustard | Mtd | GS 16948 | 69B-Gal4 |
| CG8434 | Lambik | Lbk | GS 17119 | Twi-Gal4 |
| CG7476 | Methuselah-like 7 | Mthl7 | GS 21256 | Twi-Gal4 |
| CG9342 | Microsomal triacylglycerol transfer protein | Mtp | XP d07488 | Twi-Gal4 |
| CG2005 | Protein tyrosine phosphatase 99A | Ptp99A | EY 7423 | Twi-Gal4 |
| CG13194 | Pyramus | Pyr | GS 22603 | 69B-Gal4 |
| CG8095 | Scab | Scb | EP 2591 | Twi-Gal4 |
| CG5661 | Semaphorin-5c | Sema-5c | EY 1704 | 69B-Gal4 |
| CG33950 | Terribly reduced optic lobes | Trol | GE 10067 | 69B-Gal4 |
| CG6890 | Toll-8/Tollo | | XP d01565 | 69B-Gal4 |
| CG5528 | Toll-9/Toll-like | | GS 51 | 69B-Gal4 |
| CG9138 | Uninflatable | Uif | GS 11655 | 69B-Gal4 |
| CG5993 | Unpaired/ Outstretched | Upd/Os | G17133 | Twi-Gal4 & 69B-Gal4 |
| CG10491 | Vein | Vn | GS 12044 | Twi-Gal4 & 69B-Gal4 |
| CG5372 | αPS5 | | GS 12413 | 69B-Gal4 |
| CG34056 | galactosyltransferase | | GS 11028 | 69B-Gal4 |
| CG9550 | sulfotransferase | | GS 18034 | 69B-Gal4 |

**Table 1. Twenty-four genes conferring embryonic lethality when ectopically expressed in the ectoderm or mesoderm.** Eleven genes in red indicate those that had relevant RNA expression and/or mutant mesoderm defects. FGF ligand Pyramus in blue has previously been shown to function in mesoderm migration.

CHAPTER VI

# Regulation of cadherins through FGF signaling and Snail in early *Drosophila* gastrulation

ABSTRACT

Collective cell migration requires coordination of many different components at the membrane such as signaling factors that provide directional cues and adhesion molecules to regulate cell-cell interactions. During mesoderm migration in the *Drosophila* embryo, fibroblast growth factors (FGFs) are required for several steps that support radial movement. However, it is unclear if FGF signaling is involved in guiding cells, regulating cell adhesion, or both. We investigated the relationship between cadherins, FGF signaling, and Snail (Sna), a transcription factor known to regulate E-cadherin (Ecad). Embryos mutant for cadherins have defects in mesoderm spreading, and embryos ectopically expressing FGFs show an altered distribution of cadherins. Changing the levels of FGF signaling in the embryo also led to different levels of *sna* expression, suggesting that FGFs may affect Ecad through Sna. We propose that migration of a collective is sensitive to levels of cadherins; cells unable to form contacts cannot remain in a collective while cells with excessive connections become immobile.

## INTRODUCTION

Cell-cell interactions are important for collective migration so that cells remain organized within the group and also for communicating directional information (Hammerschmidt and Wedlich, 2008). One of the earliest groups of cells migrating in the *Drosophila* embryo is the mesoderm. Mesoderm development at the onset of gastrulation comprises a migration that can be described as a multistep process. The presumptive mesoderm is first delineated during the formation of the dorsal-ventral axis, which requires the Dorsal morphogen. Dorsal activates gene expression of Twist, a transcription factor, and together Dorsal and Twist support Snail (Sna) expression in the mesoderm (Reeves and Stathopoulos, 2009). Sna is also a transcription factor and can function as an activator or a repressor of gene expression (Rembold et al., 2014). Following cellularization in the embryo, the presumptive mesoderm undergoes epithelial-to-mesenchymal transition (EMT) and invaginates to form the ventral furrow at stage 6. The furrow then collapses onto the ectoderm and the mesoderm subsequently spreads dorsally while at the same time is carried along passively by the ectoderm during germband elongation to the posterior of the embryo. Mesoderm cells finally intercalate to form a monolayer by stage 10 (McMahon et al., 2010).

Signaling by the fibroblast growth factor (FGF) pathway is known to participate in mesoderm migration. The *Drosophila* FGF receptor Heartless (Htl) is expressed in the presumptive mesoderm during cellularization at stage 5, and both Twist and Sna have positive inputs into *htl* expression (Shishido et al., 1993). The associated FGF ligands, Pyramus (Pyr) and Thisbe (Ths), are also expressed at stage 5 but in non-overlapping

domains in the ectoderm (Gryzik and Muller, 2004; Stathopoulos et al., 2004). This arrangement of receptor and ligands lends itself to a chemotaxis model where FGF signaling provides the mesoderm with directional cues to migrate dorsally.

Live imaging and tracking analysis of mesoderm migration revealed that the mechanism of FGFs is more complex than the delivery of chemoattractants. In wildtype embryos, the dorsal most cells of the furrow collapse onto the ectoderm in a radial direction and these cells migrate the farthest dorsally. Cells originating from the ventral parts of the furrow do not require large radial movement and spread only a small distance in the dorsal direction. These cells, for the most part, remain in the ventral region of the embryo throughout mesoderm migration (McMahon et al., 2008). In *fgf* mutant embryos, the dorsal-most cells of the furrow do not collapse onto the ectoderm, though they remain mobile in random directions. If these dorsal-cells come into contact with the ectoderm by chance, they acquire organized dorsal movements. Ventral-most cells of the invaginated furrow that are already neighboring the ectoderm migrate dorsally in greater distances than they normally would in wildtype to end up in lateral regions of the embryo. Cell division was unaffected in embryos lacking FGF signaling, and we detected sister cells that would sometimes migrate in different directions but eventually came back as neighboring cells. These mutants were also unable to intercalate in the radial direction at the end of the spreading process (McMahon et al., 2008). We hypothesize that these "lost" cells had defects in cell adhesion that prevented them from joining the migrating collective.

Cell adhesion through integrins and cadherins is important for cellular functions during embryonic development, including mesoderm migration (Hammerschmidt and Wedlich, 2008). We previously showed that integrins downstream of FGF signaling are

involved in cell intercalation at the end of mesoderm migration. Cells mutant for the β-integrin Myospheroid (Mys) are unable to form a monolayer (McMahon et al., 2010). Integrins are transmembrane proteins that interact with neighboring cells and the extracellular matrix (ECM). The β subunit forms a complex with α subunits, of which *Drosophila* has five α-integrins (Bulgakova et al., 2012). A recent screen identified three α-integrins potentially acting with Mys to support mesoderm migration (Trisnadi and Stathopoulos, in review). Cadherins are another family of adhesion molecules. They, along with catenins and other components, form adherens junctions that mediate actin activity along the plasma membrane (Hammerschmidt and Wedlich, 2008). One of the hallmarks of EMT, during mesoderm invagination, is the downregulation of E-cadherin (Ecad; *Drosophila* Shotgun, Shg) and the upregulation of N-cadherin (Ncad) (Baum et al., 2008). Previous studies suggest that the redistribution of E-cadherin in the mesoderm at EMT is important for spreading and is dependent on FGF signaling (Clark et al., 2011). Ecad is also dependent on Twist and Sna, both of which represses *ecad* transcription (Oda et al., 1998). This is opposite to the positive inputs Twist and Snail has on *htl* expression. While the mesoderm in *ecad* mutants are able to invaginate, spread, and later differentiate, they do not form a monolayer (Schafer et al., 2014). However, it was unclear if the multilayer defects seen in *ecad* mutant embryos were a consequence of improper EMT or if Ecad plays a more direct role in mesoderm migration. Here, we explore the relationship between Ecad, FGF signaling, and Snail with respect to mesoderm migration.

# RESULTS

*Mesoderm spreading is sensitive to FGF levels*

Embryos ectopically expressing either FGF ligands, Pyr or Ths, in the mesoderm resulted in a tight cluster of mesoderm cells and tracking analysis revealed that all mesoderm cells had minimal to no dorsal movement (Figure 1). In previous studies, *htl* mutant embryos lacking FGF signaling showed a subset of cells that migrated independently of the mesoderm collective. These cells originated from the dorsal regions of the invaginated cells at the furrow and rarely reached the ectoderm. However, cells that contacted the ectoderm, either by chance or due to their initial position within the invaginated tube, were able to spread dorsally (McMahon et al., 2008). Cell-cell interactions between mesoderm cells and also between the mesoderm and ectoderm cells appear to be important for proper migration. Collectively, these data suggested that FGF signaling modulates adhesive properties of cells to support these cell movements. We focused on cadherins, a family of adhesion molecules, which have been previously implicated in mesoderm development (Clark et al., 2011; Oda et al., 1998; Schafer et al., 2014). However, these studies rarely examined cross-sectioned embryos to determine the ability of mesoderm cells to collapse, spread laterally, or intercalate.

*Loss of cadherins results in mesoderm spreading defects*

As previously shown, the *Drosophila* cadherins have complementary expressions during mesoderm migration with Ecad in the ectoderm and Ncad in the mesoderm (Figure 2A, B). We analyzed cross-sections of embryos lacking and overexpressing cadherins and

found that, while mesoderm cells are able to reach the lateral edge, the overall layer was disordered (Figure 2C-H). In some cases of *ecad* mutant embryos, a few mesoderm cells appear within the ectoderm layer (Figure 2C, E, G). It was unclear if these cells were unable to distinguish between ectoderm and mesoderm cells. Another possibility was that cadherins provide a physical barrier to prevent cells from mixing. We therefore examined the distribution of cadherins during mesoderm migration.

*FGF signaling changes cadherin distribution*

In wildtype embryos, Ecad is initially localized in the apical sides of the ventral furrow. By the time the furrow has collapsed onto the ectoderm, Ecad has been removed from mesoderm cells (Figure 3A). However, by the end of mesoderm spreading, Ecad is concentrated between the ectoderm and mesoderm (Figure 3A, right). It is unclear if this localization of Ecad is newly synthesized or a re-distribution of Ecad from the mesoderm cells, ectoderm cells, or both. In *htl* mutants, the interface of Ecad is no longer present (Figure 3B). In embryos overexpressing either FGF ligands, Pyr or Ths, Ecad remains in the mesoderm even after collapse of the furrow. There also appears to be a seam of Ecad on the inner surface of the mesoderm tissue facing the embryo yolk (Figure 3C, D).

Upon furrow invagination, Ncad is upregulated in the mesoderm. Once cells have intercalated, Ncad forms punctate spots at the apical and basal corners that is not visible in the lateral cell surface similar to epithelial cells (Figure 3E). Ncad is still upregulated in the mesoderm in embryos lacking or increased in FGF signaling, though there appears to be a delay in *htl* mutants (Figure 3F-H). However, the localization of Ncad is no longer

restricted to the apical and basal surfaces. Punctate spots are seen throughout the mesoderm tissue, though this may be due to some indirect effect of FGFs to regulate cell shape.

*FGF signaling affects sna expression*

Sna is reported to inhibit Ecad and weakly upregulate Ncad (Oda et al., 1998). We tested the model that FGFs regulate cadherins via regulation of Sna. In wildtype embryos, *sna* is expressed in the mesoderm at cellularization and during spreading. However, at the end of mesoderm migration, *sna* switches to ectodermal expression (Figure 4A). In *htl* mutant embryos, *sna* has stronger expression in the mesoderm (Figure 4B). With ectopic expression of Pyr or Ths, *sna* has weaker expression in the mesoderm (Figure C, D). The exception is *sna* expression in the posterior-most region of the mesoderm. This area potentially overlaps with *bHLH54* expression, which later gives rise to caudal visceral mesoderm (CVM) cells and may be regulated independently.

# DISCUSSION

The mesoderm has several inputs that regulate its early development. Snail is a transcription factor important in defining the presumptive mesoderm. It activates expression of the FGF receptor and represses expression of E-cadherin (Oda et al., 1998; Shishido et al., 1993). The downregulation of Ecad is part of EMT during the invagination of the mesoderm. In subsequent steps during its migration, FGF signaling guides radial movements. We found that this process involves specific spatiotemporal distribution of Ecad and investigated the influence of FGFs on cadherins and Snail. Further analysis revealed that FGFs support E-cadherin in the mesoderm while inhibiting *sna* expression. Based on these data, we propose that FGF signaling promotes Ecad by reducing Snail levels (Figure 5). Snail is highly dynamic and appears have bursts of expression in cells just prior to their migration, such as in the CVM (Dunipace et al., 2013). It is possible that a general activity of Snail is to regulate adhesion in preparation for cell motility. However, we do not rule out the possibility that FGFs may regulate Ecad directly or through some other factor such as Twist.

Previous studies have linked these factors in other systems. Another *Drosophila* FGF receptor, Breathless (Btl), promotes Ecad levels and distribution since *btl* mutants have reduced and fragmented Ecad in the eye disc (Mukherjee et al., 2012). In vertebrate systems, FGF signaling is thought to upregulate Sna (Buxton et al., 1997). In the mouse primitive streak, ectopic Ecad was observed in *fgf* mutants along with the loss of *sna* transcript (Ciruna and Rossant, 2001). These studies suggest that there is a conserved network of these factors acting to regulate their interactions.

It is still unclear what function Ecad fulfills during early mesoderm development. One possibility is that Ecad is required for complete EMT and that improper EMT in *ecad* mutants causes mesoderm cells to have defects in invagination and/or spreading. Another consideration is that Ecad prevents ectoderm-mesoderm mixing, either as a cell-type indicator or as a barrier between layers. Electron microscopy (EM) sections revealed that during migration, the ectoderm contains multiple spot adherens junctions which is thought to be a precursor to formation of a zonula adherens belt (Tepass and Hartenstein, 1994). These membrane junctions could be necessary for monolayer formation and block further migration after intercalation.

EM also revealed that the mesoderm possesses gap junctions and spot adherens junctions (Tepass and Hartenstein, 1994). With the organized distribution of Ncad in the monolayer, it is possible that the mesoderm adopts an epithelial-like state. Other systems have shown that mesoderm reverts and undergoes mesenchymal-to-epithelia transition (MET) once migration is completed (Baum et al., 2008). Cell culture studies showed that stimulation of FGF signaling promotes MET and Ecad was ectopically expressed (Ramos et al., 2010). This would explain the random but increased movements in embryos lacking FGF activity and the immobile state when FGFs are in excess. The expression dynamics of Ecad is also consistent with this model.

Aside from mediating cell-cell adhesion, cadherins have other distinct functions that could be acting during mesoderm migration. For example, they have been shown to participate in intracellular signaling through small Rho-GTPases (Hammerschmidt and Wedlich, 2008). Previous research showed that the small GTPase Rap1 is upstream of β-integrin Mys to regulate mesoderm intercalation (McMahon et al., 2010). It is possible that

cadherins function in a similar manner and control intracellular signaling instead of or in addition to regulating adhesive properties. Another model is differential adhesion where there is a graded expression of cadherins along the collective. This predicts that cells will migrate to flank other cells with similar surface adhesiveness (Hammerschmidt and Wedlich, 2008). Recently, studies in the ovary show a shallow gradient of Ecad during border cell migration (Cai et al., 2014). The concentration of cadherins has not been quantified in the mesoderm, but future experiments will include measurements to analyze cadherin levels.

Cells require a balance between being able to coordinate with other cells within the migrating collective, but also must not be too tightly bound to each other such that they cannot migrate at all. This balance may be affected through regulation of cadherin levels and/or the distribution of cadherin molecules at the membrane. The timing of these dynamics is also important as cells must transition between Ecad and Ncad over the course of their migration. The mesoderm must also know when to begin and end their movements. Further investigations regarding the spatiotemporal regulation of cadherins will provide insight into adhesive properties of migratory collectives.

# MATERIALS AND METHODS

## *Fly strains*

Wildtype refers to *yw* or Gal4 strains: *P{GAL4-twi.G}, $w^1$* (BDSC #914) and *w; P{GawB}69B* (#1774). Mutant stocks include *$cn^1$ $shg^2$ $bw^1$ $sp^1$/CyO* (#3085), *$w^{1118}$; Mi{ET1}CadN$^{MB05059}$/CyO* (#24235), *w; $htl^{AB42}$/TM3* (#5370) and the balancer stock *$w^{1118}$; sna$^{Sco}$/CyO, P{en1}wg$^{en11}$* (#1672). *UAS-Pyr* (AMS330-3) and *UAS-Ths* (AMS289-22) were previously described in (Kadam et al., 2009).

## *In situ hybridization and immunohistochemistry*

Embryos were collected and staged at 25°C or 18°C. Antisense RNA probes were made to detect *in vivo* gene expression. Antibodies used were rat anti-Twist (1:200; housemade, Pocono), rabbit anti-β-galactosidase (1:200; Molecular Probes), rat anti-Ecadherin (1:100; DSHB), rat anti-Ncadherin (1:100; DSHB). Standard protocols for fixing and staining were used.

Cross-sections were made by embedding stained embryos in araldite (Electron Microscopy Sciences). Sections of 10-20 μm thickness were chopped using the LKB Bromma 2218 Historange Microtome and lined on a glass slide. A 1:1 mixture of acetone:araldite was added to the slide and placed on a 55°C platform to cure overnight.

## *Live imaging and tracking analysis*

Embryos expressing ubiquitous histone H2A-GFP were imaged at 940-nm two-photon microscopy. Tracking analysis was completed with Imaris software and analyzed using MATLAB. Details can be found at (Supatto et al., 2009).

FIGURES AND LEGENDS

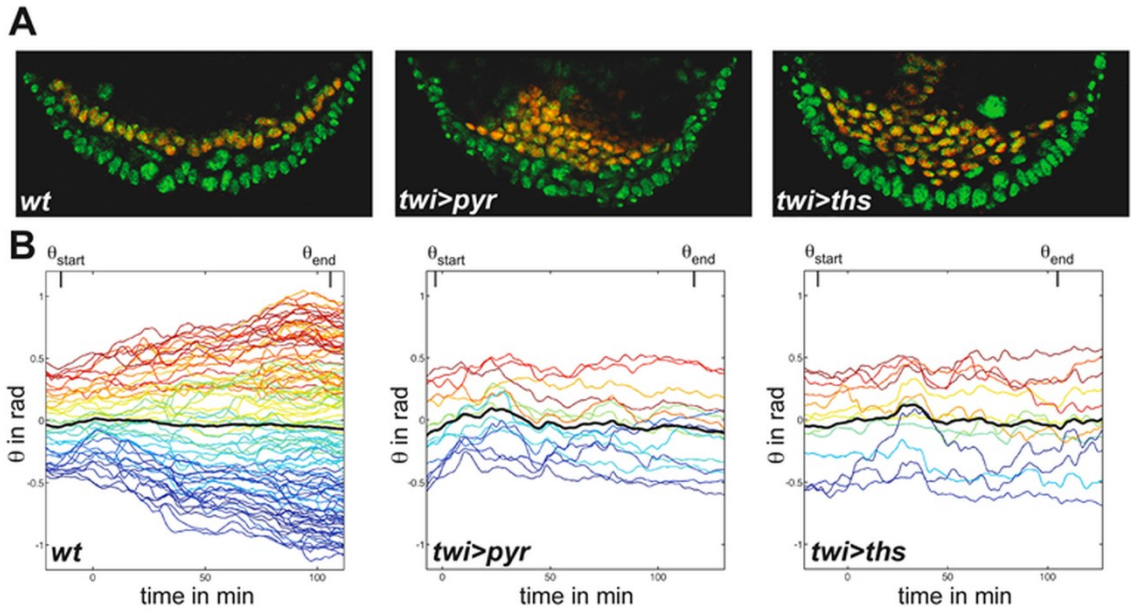



**Figure 1. Tracking analysis of embryos overexpressing FGFs.** (A) Ventral half of embryo cross-section at stage 10. Embryos overexpressing Pyr or Ths using Twi-Gal4 result in mesoderm spreading defects. Red-yellow indicates mesoderm cells; green is ectoderm. Ventral is down. (B) Tracking analysis requires the embryo to be transposed into a cylindrical coordinate system. Radial movement θ is measured in radians and represents the angle of a cell's position with respect to the ventral midline (θ = 0 radians, black line). Each line represents an individual cell and colors represent their orignial position in the ventral furrow prior to collapse. Red and blue are cells from the dorsal-lateral regions of the furrow while yellow and green are from the ventral regions plus the dorsal-most cell. Mesoderm cells with excess FGFs have little to no dorsal movement as revealed by minimal change in θ throughout the course of the migration. Data from A. McMahon.

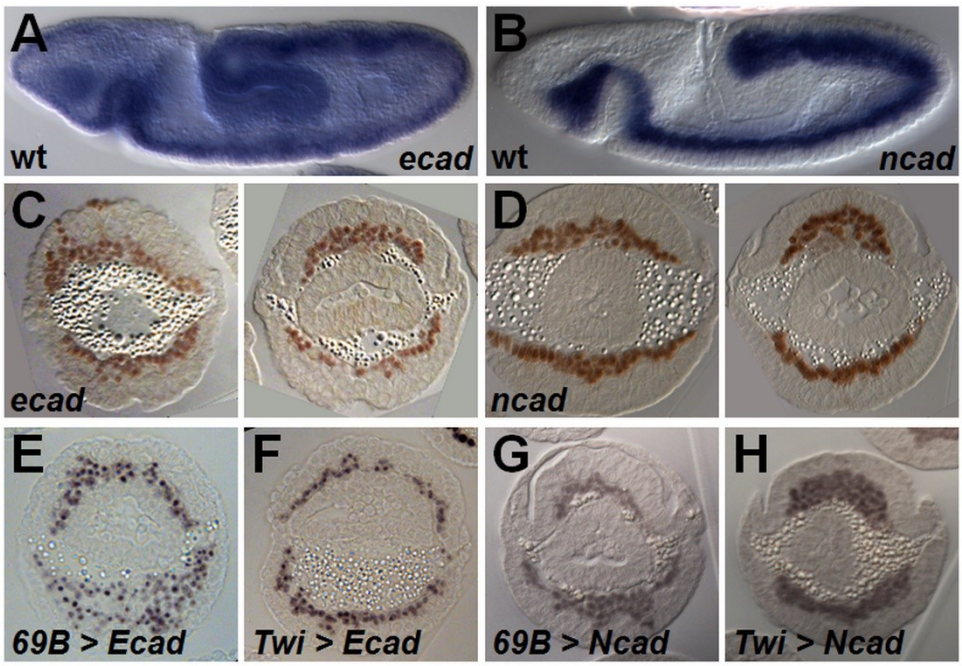



**Figure 2. Characterization of cadherins in mesoderm development.** Expression of (A) *ecad* and (B) *ncad* in wildtype embryos. Cross-sections of embryos lacking (C) Ecad or (D) Ncad. Cross-sections of embryos overexpressing Ecad in the (E) ectoderm or (F) mesoderm. Overexpression of Ncad in the (G) ectoderm or (H) mesoderm. Embryos are stage 10; anterior is left, ventral is down.

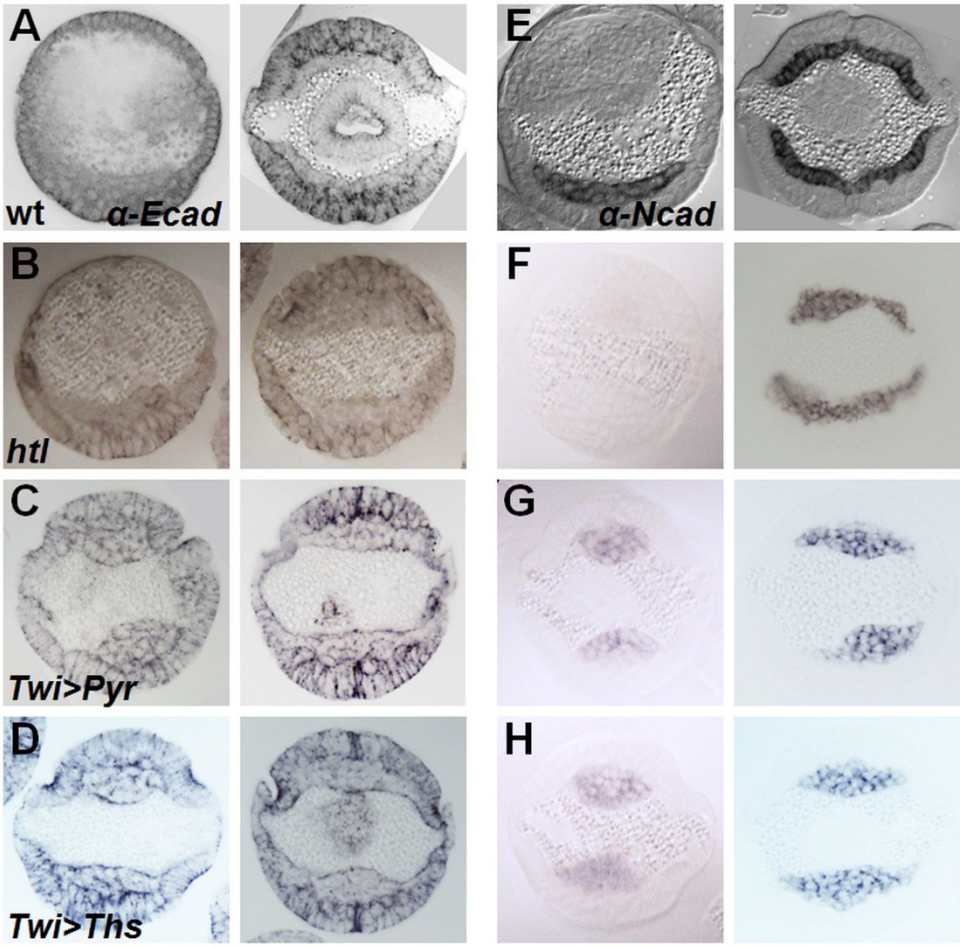



**Figure 3. Distribution of cadherins during mesoderm migration.** Embryo cross-sections of Ecad immunostaining in (A) wildtype, (B) *htl*, (C) Pyr overexpression, and (D) Ths overexpression embryos. Ncad immunostaining in (E) wildtype, (F) *htl*, (G) Pyr overexpression, and (H) Ths overexpression embryos. FGFs were overexpressed in the mesoderm using Twi-Gal4. Left columns are stage 8, right is stage 10. Ventral is down.

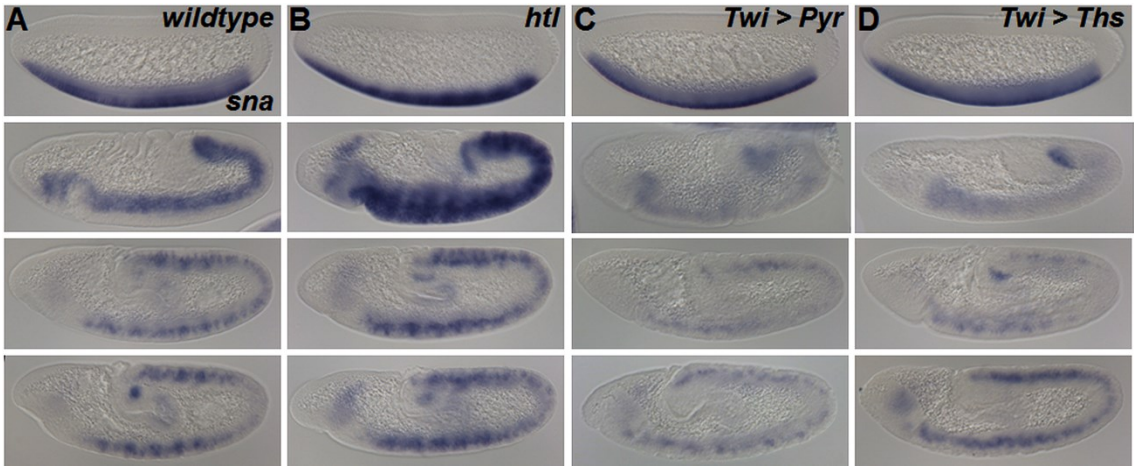



**Figure 4. Expression of *sna* in *fgf* mutant embryos.** Comparison of *sna* expression in (A) wildtype, (B) *htl*, (C) *Twi>Pyr*, and (D) *Twi>Ths* embryos. From top to bottom, embryos are aged at stage 5 cellularization, stage 7, stage 9, and stage 11. Anterior is left, ventral is down.

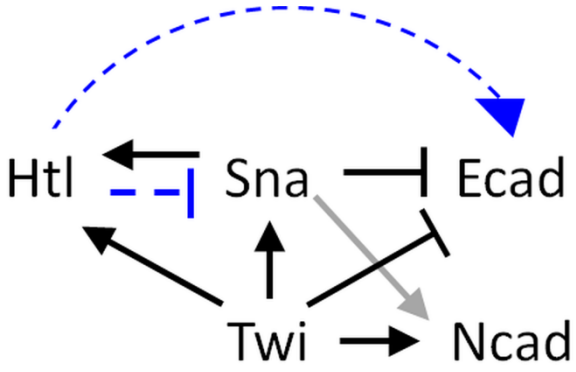



**Figure 5. Proposed network regulating mesoderm adhesion.** Previous studies show that Sna promotes *htl* expression while repressing *ecad* expression. Sna also weakly supports expression of *ncad* (gray arrow). Twi also has positive inputs in *htl*, *sna*, and *ncad* expression but negatively regulates expression of *ecad*. We suggest a model where Ecad requires FGF signaling, which may be a direct input or indirect through Sna (blue dashed lines).

CHAPTER VII

# Discussion

**Patterning with gene expressions that are spatiotemporally dynamic**

Gastrulation requires patterning to delineate specific domains of gene expression. While traditional studies of patterning have mapped the connections and interactions that give rise to precise expression domains, these gene regulatory networks do not always cover the spatiotemporal dynamics of gene expression. During embryonic development, there are many changes in the environment surrounding the cell and these changes can occur on minute timescales. This work investigates the influence of the spatiotemporal dynamics in gene expression during dorsal-ventral patterning and mesoderm migration in the *Drosophila* embryo.

*Regulation of Dorsal-mediated gene patterning*

The Dorsal morphogen gradient patterns the dorsal-ventral (DV) axis during a time when nuclei are rapidly changing and the lack of cellular membranes limits spatial regulation. It was not clear how Dorsal is able to establish gene expression domains in this dynamic environment. The extent to which nuclei can interpret the gradient in regions of low morphogen concentration was also unknown. Two-photon light-sheet microscopy (2P-SPIM) had the necessary spatiotemporal resolution that allowed us to characterize Dorsal dynamics in real time. We found that throughout nuclear cycles 11-14, Dorsal levels increased in ventral nuclei and decreased in dorsal nuclei. Our measurements indicate that the Dorsal gradient does not reach steady state, unlike the Bicoid morphogen which patterns the anterior-posterior axis (AP) and remains at peak levels starting in nuclear cycle (nc) 10 (Gregor et al., 2007b). While Bicoid is known to be locally translated, it is yet to be

determined if Dorsal is continually being synthesized in a similar manner. Another possibility is that there is a pool of Dorsal protein which flows from dorsal to ventral regions, either by diffusion or active transport. Fluorescence recovery after photobleaching (FRAP) experiments with Dorsal-GFP would help resolve this ambiguity (Figure 1). Our discovery of a spatiotemporally dynamic Dorsal gradient led us to question how downstream target genes were able to reliably pattern the embryo.

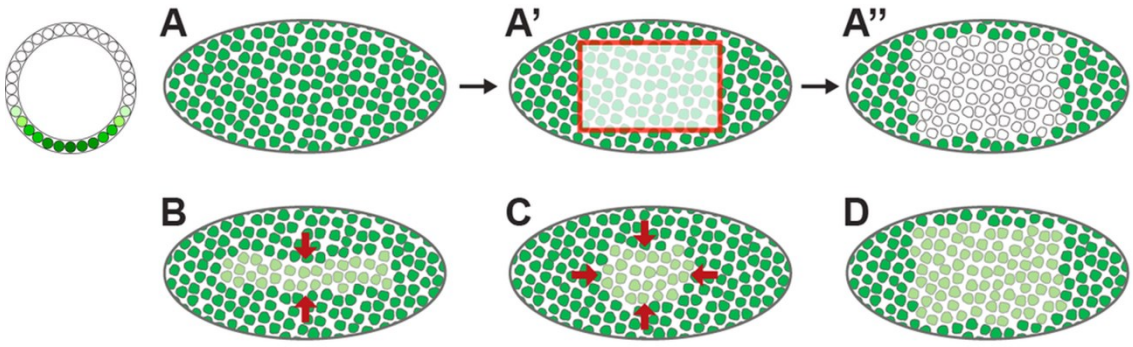



**Figure 1. Predicted mechanisms of Dorsal dynamics using FRAP analysis.** (A) Schematic of the ventral view of the *Drosophila* embryo with Dorsal-GFP. (A') An area in the ventral region is bleached at the onset of nc 14. (A'') GFP signal inside the area is removed. (B) If there is a flow of Dorsal from dorsal into ventral regions, the area will be filled with Dorsal-GFP laterally. (C) It is also possible that Dorsal-GFP enters from all directions. (D) Newly synthesized Dorsal-GFP can also explain the increase of Dorsal in ventral regions.

We analyzed the expression domains of Type I-III Dorsal target genes over nuclear cycles 11-14. Embryos were carefully staged and the longer nuclear cycles 13 and 14 were further divided to capture subtle shifts in boundaries. We found that that these genes, such as *sna* and *sog*, were dynamic as well and that their changes correlated with the Dorsal gradient. Shifts in the borders, as well as repressive activity, can be observed within nc 13 and nc 14. For example, the *sna* domain expands slightly in nc 14 which can attributed to

the Dorsal increase in ventral regions. At the same time, Sna repression of *sog* is absent at the onset of nc 14 and is only detected later in nc 14. This is the first time the spatiotemporal dynamics of Dorsal target genes are observed over a period of multiple nuclear cycles, though similar findings with Bicoid downstream genes have been reported. While the Bicoid gradient was not observed to be dynamic, many of its target genes regulate each other. These cross-repressive activities are likewise delayed due to the design of the network where Bicoid must first activate the initial set of gene expression (Jaeger et al., 2004). Despite the differences in Dorsal and Bicoid dynamics, both are able to use these common themes in network interactions and reliably establish their respective axis in the embryo. Collectively, these findings emphasize that gene patterning can robustly take place in the face of spatiotemporal dynamics of the environment and multiple regulatory inputs.

In order to evaluate locations of these gene borders with respect to the embryo axis, we developed a protocol for quantifying domains of gene and protein expression. The exact concentration levels of transcripts cannot be calculated or compared between genes using this method. However, new techniques such as qPCR and NanoString have also given us the ability to quantify the amount of mRNA present within a single embryo (Kulkarni, 2011). We are able, though, to compare embryos of varying sizes and, for example, to ask questions regarding the scaling of genes along the DV axis. Patterns are said to scale if their expression domains change relative to the size of an embryo. Using our method, the width of the Dorsal gradient was measured and found to scale according to the length of the DV axis. This scaling extended to many Dorsal target genes, but not all such as *ind* in the neurogenic ectoderm (Garcia et al., 2013b). This indicates that *ind* may have independent inputs to regulate its expression. Since scaling involves refinement of expression borders,

which we have shown can shift over time, it is necessary that embryos are staged precisely to ensure that scaling dynamics have reached their final output.

Future investigations using our protocol would continue determining spatiotemporal dynamics of expressions in additional Type I-III genes and include mutant analysis. The rapid time intervals certainly appear to play an important role. Recent characterization of the RNA Polymerase II mutant X161 prolongs the length of nc 13 and eliminates nc 14 (Sung et al., 2013). However, it remains unclear how patterning is affected by changes to the duration of nuclear cycles in these mutants. It is possible that prolonging the length of nuclear cycles will allow Dorsal to continue increasing its concentration in ventral regions without the interruption of a mitotic reset. As a consequence, boundaries of downstream target genes may be affected, for example leading to their overexpansion, and result in improper patterning. Gene expression domains in other systems can also be examined. Additional applications of this protocol can extend beyond *Drosophila* embryos since these analysis do not require that inputted image be a circular cross-section.

Not all gene domains have sharp boundaries, and several mechanisms are predicted to generate graded borders. Integrating noise is one possible method to control expression. Our simulations indicate that a certain amount of noise is required to support patterning that closely matches observed borders. Furthermore, modeling revealed that the spatiotemporal dynamics of gene expression must also be an input in order for the noise to be effective. In addition, we found that gene length is another potential mode of regulation for graded boundaries (McHale et al., 2011). This mechanism can be implemented through multiple introns to increase transcript length and/or alternative splicing. Perhaps undiscovered, shorter spliceforms exist in the embryo and are functional only during early development.

Patterning in the early embryo requires multiple inputs at many levels to yield precise domains of gene expression. Changes occur over a landscape with increasing number of nuclei and in minute timescales. These spatiotemporal dynamics of the cell continue as the embryo develops, and we find that additional tiers of regulation are a common theme to controlling gene patterning.

*Spatiotemporal expressions support early mesoderm development*

Once the embryo has been patterned through differential gene expression, it begins to gastrulate into separate germ layers. We focused on early development of the mesoderm. Briefly, the presumptive mesoderm begins epithelial-to-mesenchymal transition (EMT) and invaginates to form an internalized tube of mesoderm cells. This tube subsequently collapses onto the ectoderm and spreads dorsally. The final step ends with intercalation to form a monolayer of mesoderm cells. Previous investigations characterized the role of Fibroblast growth factor (FGF) signaling during mesoderm migration. It was discovered that FGFs are required in a subset of mesoderm cells for its radial movements, which includes furrow collapse and intercalation (McMahon et al., 2010; McMahon et al., 2008).

Further analysis revealed that the spatiotemporal expressions of two heparan sulfate proteoglycans (HSPGs) are important in regulating FGF signaling (Trisnadi and Stathopoulos, in review). The HSPG Terribly reduced optic lobes (Trol) is secreted while Syndecan (Sdc) is membrane-tethered. We found that these two HSPGs are both expressed in the ventral regions of the ectoderm. However, only Trol is present earlier in the migration when the mesoderm furrow collapses onto the ectoderm. Sdc, on the other hand, is expressed later in the process during intercalation. The timing of their expression

coincides with the stages when FGF signaling is guiding the mesoderm in radial movements. We believe that this switch from Trol to Sdc, considering their localization, supports modulation of FGF distance of action. Trol facilitates FGF signaling during long-range communication and Sdc is utilized in short-range FGF signaling. We show other instances relating these two HSPGs in FGF-dependent processes that support our model.

One possible reason for the lack of radial movements in *fgf* mutants is the disruption in cell adhesion. Previous work has implicated cadherins and integrins in mesoderm migration, both of which are responsible for adhesion (Clark et al., 2011; McMahon et al., 2010). However, during EMT in the mesoderm, E-cadherin (Ecad) is downregulated and N-cadherin (Ncad) is upregulated (Baum et al., 2008). It was unclear if phenotypes seen in cadherin mutants are a consequence of improper EMT or if they have a direct role in mesoderm migration. We examined the spatiotemporal distribution of E-cadherin (Ecad) and found defects in its localization when FGF signaling levels were altered. In addition, we discovered an indirect interaction between the FGF pathway and *ecad* expression. We showed that FGF signaling inhibits *sna* expression, and Sna is a known repressor of Ecad (Oda et al., 1998). Therefore, FGFs could be promoting Ecad through Sna. It is also possible that FGFs can activate *ecad* expression directly, but further studies are required to confirm this link. With the spatiotemporal complexity of cadherin gene expression and protein distribution, as well as the requirement of integrins, we suggest that a balance of adhesive properties is important in a migrating collective.

*Elusive factors directing the mesoderm*

The FGF pathway is responsible for mesoderm migration in the radial direction, but it remains unknown what is guiding cells to spread dorsally. We conducted a screen of membrane-bound and secreted factors in hopes of identifying the pathway(s) and/or cell-cell contacts responsible for dorsal movement. Several classes of genes were isolated, including signaling components, modifying enzymes, proteoglycans, and adhesion molecules. Further research is required to determine their exact roles in mesoderm migration, but initial analysis did not distinguish a potential function in dorsal spreading. It is possible that ectopic expression in restricted spatial domains was not sufficient to induce lethality, our assay during the first selection and a probable shortcoming of the screen design. It is also possible that we would not detect any spreading defects if FGF signaling is adequate. Another survey in an *fgf* mutant background would resolve this ambiguity.

Several mechanisms that do not necessarily employ signaling pathways can be considered (Figure 2). One is a simple space-filling model in which the mesoderm cells will spread to occupy all available areas. Another method could include communication with neighboring cells, the ectoderm substratum and/or the ECM, through gap junctions or other membrane components. Finally, differential adhesion could drive cells to migrate towards regions with similar surface affinity. This would require membrane factors, such as integrins and/or cadherins, to be expressed in a graded fashion (Hammerschmidt and Wedlich, 2008).

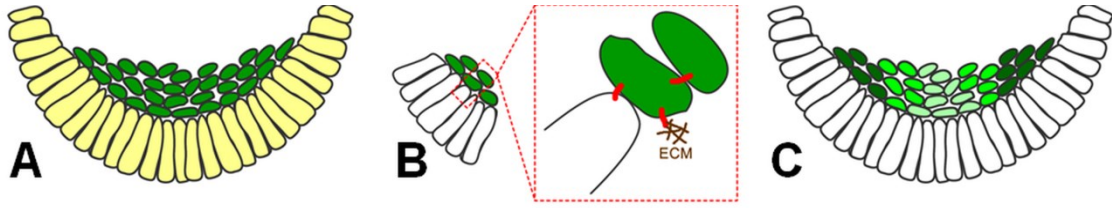



**Figure 2. Models of guiding mesoderm lateral spreading using surface interactions.** (A) Space filling models predict that cells invade to all open areas. (B) Surface communication with ectoderm substratum through gap junctions and/or with ECM components to direct spreading. (C) Cells with differential adhesion will reorganize to contact those with similar properties.

Another area of interest is *in vivo* imaging of (1) FGF receptor and ligand proteins and (2) the cellular membranes during migration. Two ligands function in mesoderm migration, but their specificity is unclear. The dynamics of ligand diffusion, not to mention the FGF-FGFR complex, would give insights into FGF signaling properties. There is a plethora of information regarding the dynamics of protrusions and cell motility *in vitro*. Unfortunately, intrinsic cues are not always known and cells are sensitive to external cues. Therefore *in vitro* experiments will never recapitulate the interaction between neighboring cells, friction from the ECM, and other environmental features. Still, there are challenges that prevent real-time, *in vivo* imaging. Protrusions are thin and push the limits of imaging resolution. The membrane dynamics of migrating cells are often fast and traditional microscopy is unable to capture these changes at the speed required to obtain decent signal, especially when it is deep within the specimen. However, new imaging technologies will soon allow us to visualize membrane dynamics during migration *in vivo*.

*Seeing is believing*

Gastrulation of an embryo requires integration of multiple processes. Recent advances in imaging technology allow us to address questions from many levels. New studies are able to visualize replication and transcription in real time during early embryogenesis (Garcia et al., 2013a; Lucas et al., 2013; Yuan et al., 2014). There have also been developments of photoactivated reagents that have become available (Cai et al., 2014). This has allowed us to address questions of protein function in addition to gene expression. This includes protein activation, proximity sensors, and modifications to the protein and genome. For example, the migration of border cells can be manipulated *in vivo* using photoactivatable Rac. This has led to our ability to probe the environment and identify regions where cells cannot travel (Wang et al., 2010).

Studies into collective migration may benefit from tracking analysis, and when the system is comprised of a thousand cells, individual labeling greatly facilitates the process. To this end, photoactivated and photoconvertible fluorescent proteins can be applied. We have developed a system using Dendra (Evrogen), a green-to-red fluorescent protein that has been fused to histone H2A to mark nuclei and integrated into the *Drosophila* genome under a ubiquitous promoter (Figure 3). Lifeact is another marker that binds F-actin and is a promising candidate for visualizing cellular membranes (Hatan et al., 2011). As our imaging capabilities push the limits in spatiotemporal resolution, we can observe dynamics in real time *in vivo*.

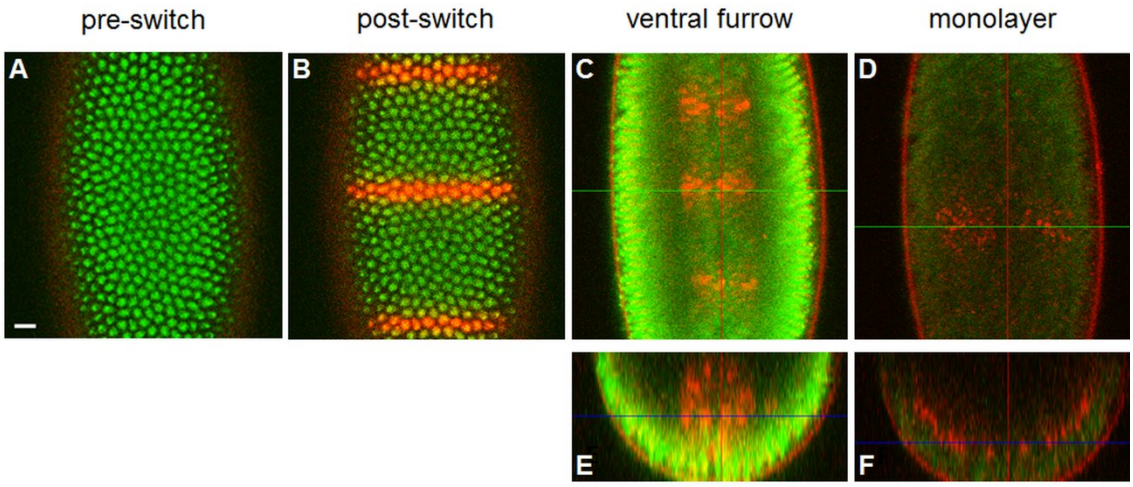



**Figure 3. Dendra2, a green-to-red photoswitchable protein, expressed in the *Drosophila* embryo.** Live imaging of a wildtype embryo expressing $Ubi_{prom}$-Histone-Dendra2. (A) Stage 5 embryo at the end of nc 14. (B) Three rows of cells were photoswitched using a 405-nm laser. (C, E) Embryo has developed to stage 6 when the ventral furrow is formed. (D, F) Embryo is now at stage 10 when cells have completed intercalation to form monolayer. Only one row of switched cells remains in ventral regions. Other rows are on the dorsal side, carried by the ectoderm during germband elongation. (A-D) Ventral views, anterior is up. (E, F) Optical cross section, ventral is down. Green lines in C and D indicate location of optical cross-sections in E and F. 10μm scale bar.

Advances in imaging technology have allowed us to investigate many of the dynamics inherent in embryonic development and appreciate the changes that occur within minute timescales. Genes are activated and immediately the transcripts are degraded during mitosis. Initial boundaries of gene domains can shift and refine over one nuclear cycle. The spatial restrictions and temporal requirements of gene expression influence signaling pathways. These observations introduce layers of complexity on top of the gene regulatory networks in patterning. We conclude that the spatiotemporal dynamics found in the environment and regulatory contributions are important to support gene patterning and orchestrate gastrulation.

# Appendices

APPENDICES

# A. Supplementary Materials for Chapter II

***Dorsal-Ventral Gene Expression in the Drosophila Embryo Reflects the Dynamics and Precision of the Dorsal Nuclear Gradient***

SUPPLEMENTAL INVENTORY

**Supplemental Figure and Movie Legends:**



**Supplemental Experimental Procedures:**

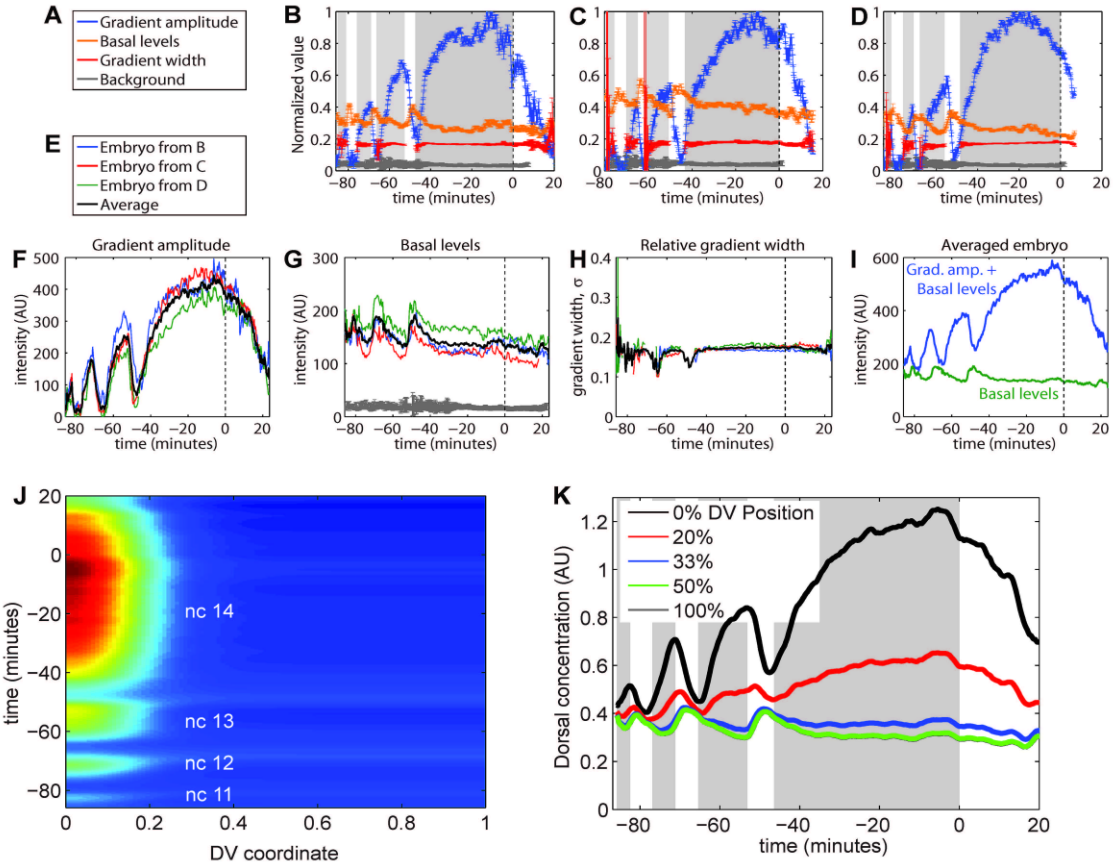



**Figure S1: Measurements and averaging of three live embryos. Related to Figure 3.** (A) Legend for plots in B-D. (B-D) Plots of the gradient amplitude, basal levels, and gradient width (interphase only) of Dorsal-Venus from three separate live embryos. Embryo in B was analyzed for Figure 3D-I. The gray curve at the bottom represents the background levels, which is the intensity of the Venus channel in a control embryo carrying H2A-RFP only. The background levels should be compared to the basal levels and not to the gradient amplitude or width. Errorbars denote 68% confidence intervals on the fitted parameters. (E) Legend for plots in F-H. (F) Plot of gradient amplitudes of the three embryos with the durations of the interphases and mitoses aligned. The black curve represents the average of the three embryos. (G) Same as F except with basal levels. The

gray curve at the bottom is the background levels from the control embryo. (H) Same as F except with the gradient width. (I) Averaged basal levels (green) and gradient amplitude plus basal levels (blue). (J) Heatmap of Dorsal nuclear levels over time and space averaged from three live embryos. (K) Traces of averaged Dorsal concentration seen by nuclei at five different DV locations.

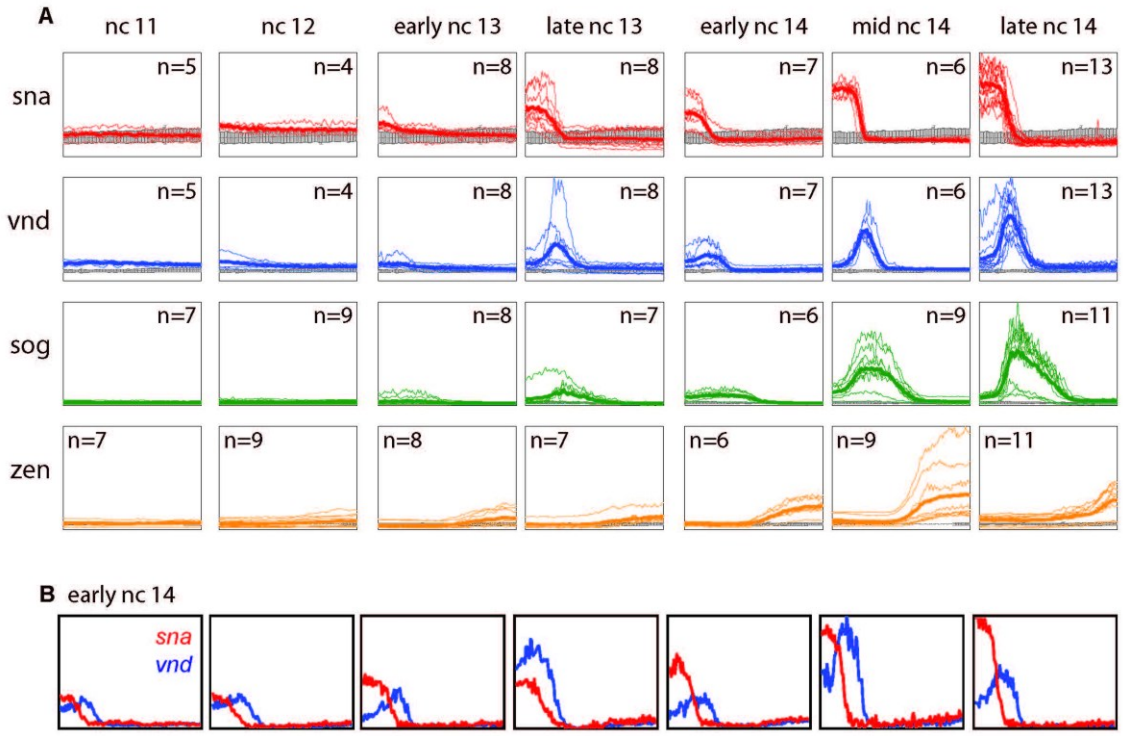



**Figure S2: Profile overview of Dorsal target genes. Related to Figure 4.** (A) Overlay of individual profiles for each Dorsal target gene at each nuclear cycle, with the thicker line representing the average and *n* being the number of embryos analyzed. Gray curves represent background levels specific to the mRNA antibody and channel (errorbars standard deviation; see Supplemental Experimental Procedures). (B) Profiles of individual embryos at early nc 14 co-stained with *sna* (red) and *vnd* (blue) show a large range in intensity and pattern even within a nuclear cycle substage.

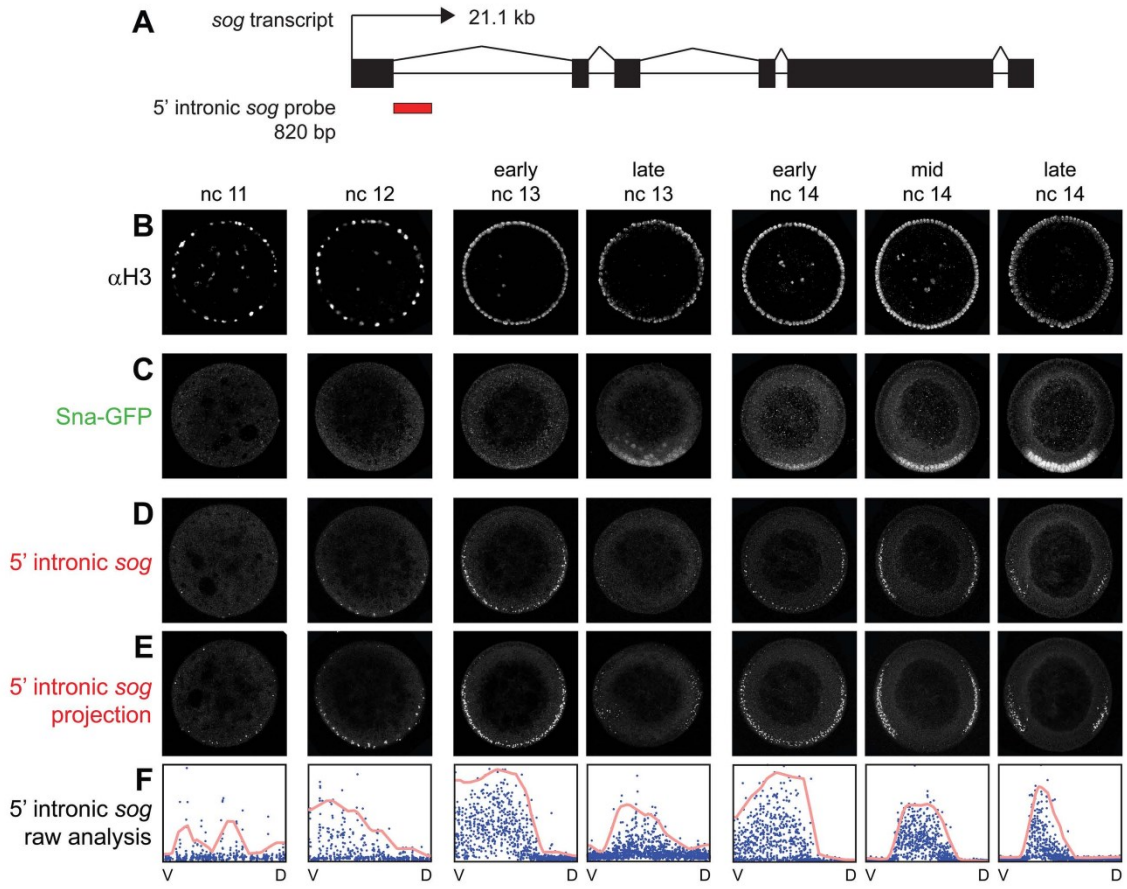



**Figure S3: Cross-sections of embryos with Sna-GFP and intronic *sog*. Related to Figure 5.** (A) The 21.1 kb *sog* transcript contains 6 exons and 5 introns. The 820 bp intronic *sog* probe used in this study starts at the beginning of the first intron (red bar). (B) 1.3 micron optical slices of embryos containing the Sna-GFP transgene are shown stained with nuclear Histone H3. (C) The same embryo for each stage was also stained with anti-GFP for the Sna-GFP protein. (D) Same embryo from B and C except with intronic *sog*. (E) The 19.5 micron z-stack projection is displayed to fully capture the intronic *sog* expression. (F) Raw analysis of the single embryos shown in B-E with blue dots representing intronic *sog* and its corresponding profile curve in pink. Analyses of additional embryos are shown in Figure 5G-J. D: dorsal, V: ventral.

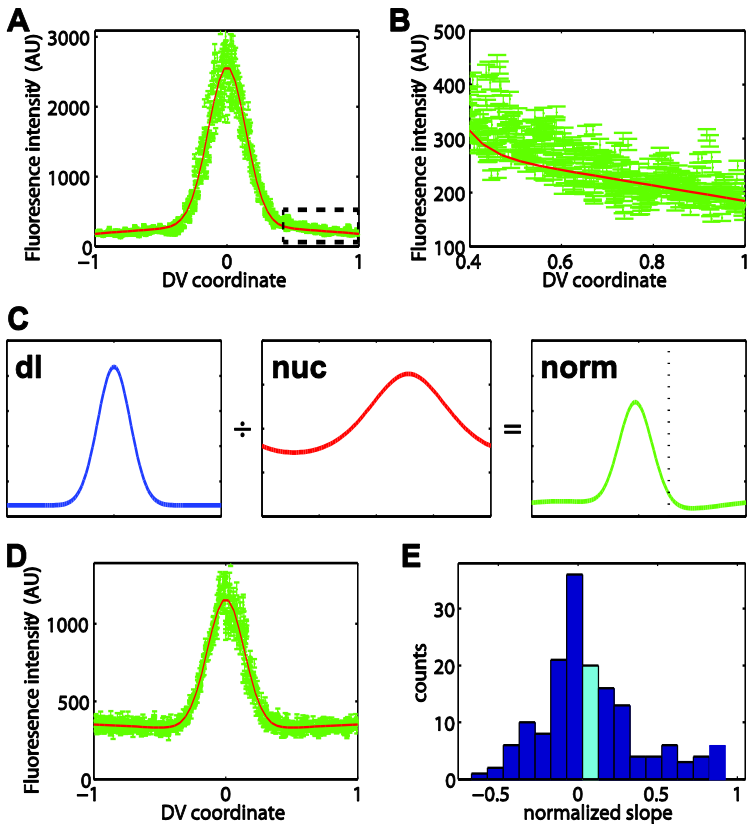



**Figure S4: Detecting the slope of the gradient tail. Related to Figure 6.** (A) Dorsal nuclear gradient for a representative embryo (with normalized gradient tail slope = -0.06). Outside the Gaussian regime, the tail appears to slope gradually downward in a shallow linear fashion. Dashed box represents the portion of this plot that is depicted in part B. Errorbars denote the standard error of the intensity of the pixels in each nucleus (also in B,D). (B) Same embryo from A, but zoomed-in on the tail. While the tail could be described by other functions, it is shallow enough such that a one-term Taylor expansion is sufficient. (C) Scheme of numerical controls to show that the slope of the gradient tail is, on average, negative. The blue curve is a hypothetical Dorsal gradient assuming the tail is flat. The red curve is a possible non-uniformity in the intensity of the nuclei, based on real

images. The peak of this curve has been randomly placed with respect to the peak in the Dorsal gradient (i.e., the presumptive ventral midline). When the Dorsal gradient is normalized by the nuclear intensity (green curve), artificial x-dependence emerges. Dotted line: random placement of the peak of nuclear intensity. (D) Embryo (real data) in which gradient tail slope is positive. (E) Histogram of simulated gradient tail slopes. The mean is 0.08 with a 95% confidence interval of the mean of [0.03, 0.13].

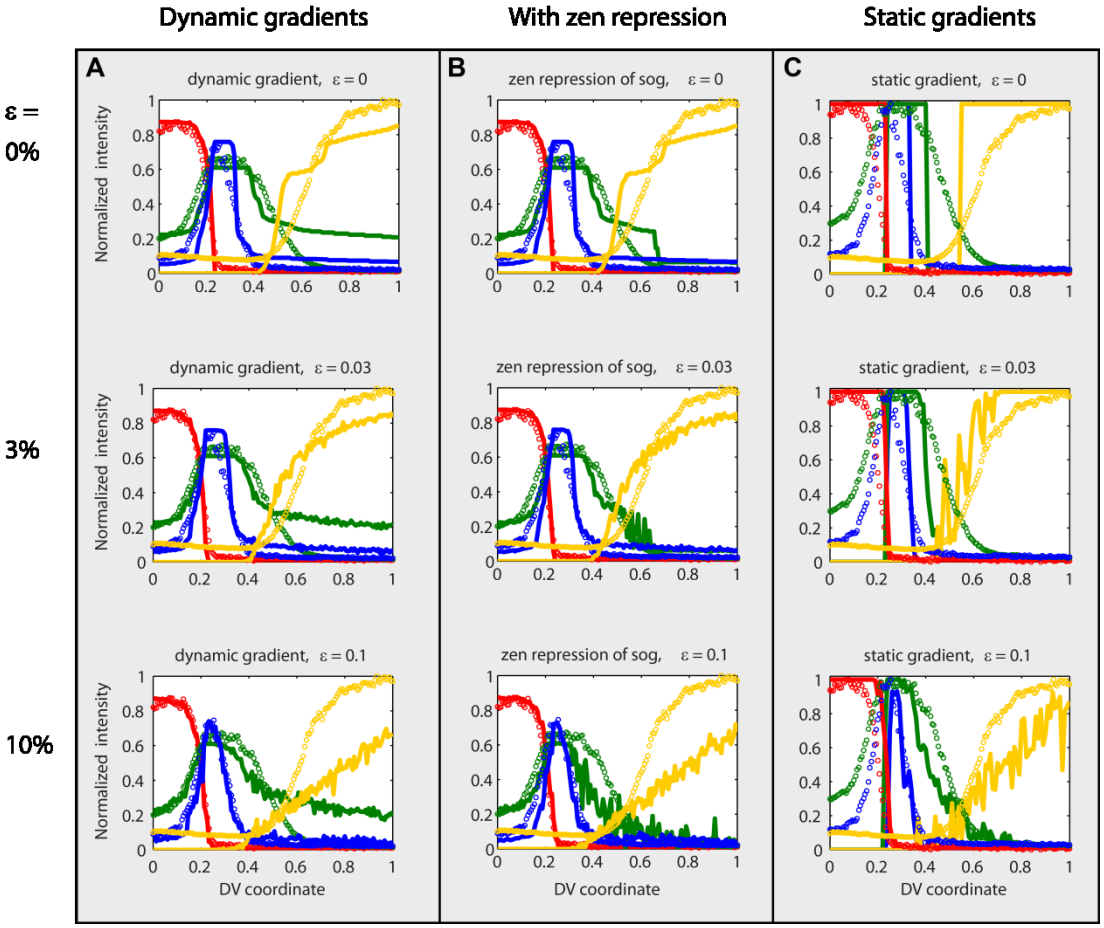
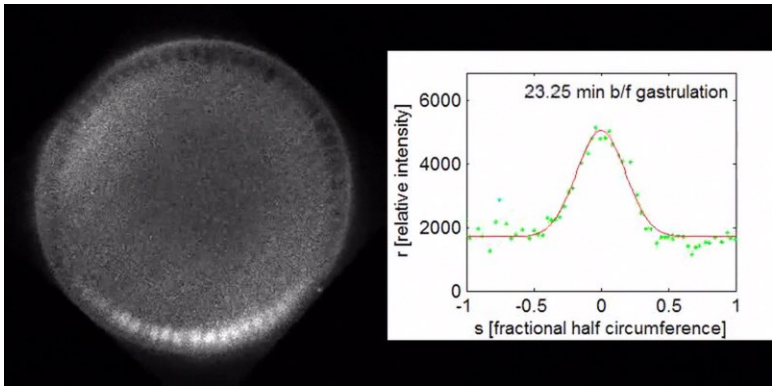



**Figure S5: Simulations of mRNA patterns resulting from dynamic and static Dorsal gradients and different levels of stochastic noise. Related to Figure 7.** In this figure, the noise level, ϵ, was varied to determine the effect this parameter had on the mRNA patterns. (A) In dynamic gradients, little effect was found for the Type I and II genes, but the graded response of the Type III genes was affected. In general, too much noise was adverse. However, even without noise, the Type III genes maintained graded borders. This demonstrates that the basal levels decreasing may be one factor that contributes to graded mRNA patterns. (B) When repression of *sog* by *zen* is included, the graded dorsal expression of *sog* better matches observed patterns indicating the possibility of a repressor.

(C) In static gradients, some small amount of noise is essential for creating graded patterns, but too much noise was adverse. This demonstrates that a noisy gradient tail and time-averaging may be one factor that contributes to the graded patterns of Type III genes. Shown are the late nc 14 Dorsal target genes *sna* (red), *vnd* (blue), *sog* (green), and *zen* (yellow). Circles denote averages of fluorescent in situ hybridization patterns from > 10 embryos, and solid curves denote simulation results.




**Movie S1. Live imaging of Dorsal-Venus using two-photon light-sheet microscopy. Related to Figure 3**. The Dorsal-Venus signal from *dl*$^{1}$*/+; dl-venus/H2A-RFP* mothers is presented on the left. The embryo was imaged at 150 microns from the anterior pole starting at nc 11 until the beginning of gastrulation. The accompanying plot on the right displays analysis corresponding to each timepoint of the movie. Intensity level of each nucleus at the given location along the DV axis is represented by green dots and the fitted Dorsal gradient is shown by the red curve.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

**Primers used**

*Dorsal homology arms*
dl LA-AscI-F: AGGCGCGCCCGCTGCTGATATGATGGTTG
dl-LA-BamHI-R: CGCGGATCCGATTTGTCCAGAAACCTGTG
dl-RA-BamHI-F: CGAGGTAATTTTTAATGGATCCTGCC
dl-RA-AsiSI-R: AAGGAAAAAAGCGATCGCCTGGAACTGTGTCTTTATC
*GalK primers*
Dorsal-GalK-F: TGCGC CTC AAT TCG GAA GAT CTG CAG ATA TCG AAC CTG TCC ATA TCC ACG GAA GGA GGC GGT GGG GGT CCT GTT GAC AAT TAA TCA TCG GCA
Dorsal-GalK-R: CT ACT GAC TCC TCC GTT CTT GCT CTG CTC TGG TTC GTT GTG AAA AAG GTA TCA GCA CTG TCC TGC TCC TT
*Venus insertion and adding 6XGly*
Dorsal-Venus-F: TGCGC CTC AAT TCG GAA GAT CTG CAG ATA TCG AAC CTG TCC ATA TCC ACG GAA GGA GGC GGT GGG GGTATGGTGAGCAAGGGCGAGGA
Dorsal-Venus-R:
CTACTGACTCCTCCGTTCTTGCTCTGCTCTGGTTCGTTGTGAAAAAGGTA
CTTGTACAGCTCGTCCATGCCC
*GFP insertion and adding 6XGly*
dl-GFP-F: TGCGC CTC AAT TCG GAA GAT CTG CAG ATA TCG AAC CTG TCC ATA TCC ACG GAA GGA GGC GGT GGG GGT ATGAGCAAGGGCGAGGAACT
*GFP insertion with SV40 terminator*
dl6xglyGFP-f:
CTCAATTCGGAAGATCTGCAGATATCGAACCTGTCCATATCCACGGAAGGAGGC
GGTGGGGGTATGAGCAAGGGCGAGGAACT
dlGFPkan-r:
TCATATCATCATCCTACTGACTCCTCCGTTCTTGCTCTGCTCTGGTTCGTTGTGAA
AAAGGTATCGAAGAGCTATTCCAGAAGTAGTGA

*Dorsal-fluorescent protein fusions*

Two slightly different *dorsal-GFP* constructs were used in this study. The *dorsal-GFP* construct used for live *in vivo* imaging was cloned analogous to *dorsal-venus*, except this first *dorsal-GFP* created contains an additional terminator sequence, SV40, following the *gfp* gene. Therefore, the construct was remade as a seamless insertion of *gfp* into the

*dorsal* locus (exactly as *dorsal-venus* was constructed) to produce a seamless version: *dorsal-gfp* (seamless). Nevertheless, fixed analysis of *dorsal-GFP* (seamless) showed that even this construct supported a dorsal gradient that was wider than that supported by *dorsal-venus*. Thus, the wider gradient associated with *dorsal-gfp* is not due to the SV40 terminator sequence. Moreover, both *dorsal-GFP* (seamless) and *dorsal-GFP* (which contains the SV40 terminator sequence) require two copies to rescue the *dl* mutant, unlike *dorsal-venus*, which complements at one copy. These GFP constructs have significantly larger widths compared to wildtype and Dorsal-Venus [Figure 1F, *dl-GFP* (live) and *dl-GFP* seamless (fixed)].

In addition, the *dorsal-venus* construct contains the following sequence from its 3'UTR before reaching the stop codon: YLFHNEPEQSKNGGVSRMMI.

*Sna-GFP fly stock*

Analysis of the Snail protein was done through antibody staining of GFP in transgenic embryos containing a 25 kb Sna-GFP rescue transgene previously described (Dunipace et al., 2011). This construct includes the endogenous 3' UTR as well as an SV40 terminator sequence associated with the GFP insertion, and importantly is able to complement *sna* mutants. More information can be found in Dunipace et al., 2011.

*Embryo antibody stainings*

We performed double in situ and antibody fluorescent stainings using standard protocols but eliminated Proteinase K treatment (Kosman et al., 2004). Antisense RNA probes were made against *sna, vnd, sog, 5' intronic sog, zen, ths,* and *Neu3*. Primary

antibodies used are anti-Dorsal 7A4 monoclonal mouse (DSHB), anti-GFP polyclonal goat (Rockland 600101215) for GFP and Venus detection, anti-Histone H3 polyclonal rabbit (Abcam 1791), anti-DIG mouse (Roche 11333062910), anti-FITC goat (Rockland 600101096), and anti-BIO goat (Rockland 600101098). Secondary antibodies from Invitrogen used are Alexa Fluor 488 anti- goat (11055), Alexa Fluor 555 anti- goat (21432) and mouse (31570), and Alexa Fluor 647 anti- rabbit (21245).

*Image analysis*

For optical sections of embryos, the perimeter of the embryo was found based on the local drop in intensity in the radial direction for 60 points equally-spaced in the azimuthal angle, similar to the method described in Liberman et al., 2009. To detect gene expression, first 300 equally-spaced points were placed around the perimeter of the embryo, interpolating from the original 60 points. Second, a series of quadrilaterals was defined by two adjacent points on the perimeter and two corresponding points 20 microns closer to the center of the embryo. The intensity of gene expression at each point around the perimeter of the embryo was computed as the mean fluorescence intensity inside each quadrilateral.

Nuclei were detected in the following manner. First, the nuclear layer was unrolled to 20 microns deep into the embryo, transforming the annular nuclear layer into a strip, as described previously (Liberman et al., 2009). The fluorescent intensity was averaged along the radial axis of the embryo to give a 1D approximation to the nuclear layer. This 1D approximation was morphologically opened using a line of width 3 microns, and boundaries between adjacent nuclei were determined based on a watershed algorithm.

Rectangles in the strip of nuclei were defined by the locations of these boundaries, and within each rectangle the raw nucleus was segmented using a best-fit threshold protocol (Otsu, 1979). To eliminate spurs and feathers, each raw nucleus was morphologically opened using a disk of radius 1.5 microns, yielding a final set of nuclei in the strip of the nuclear layer. The locations of each "on" pixel in the strip were then transformed back into the original 2D embryo image, maintaining the distinction between neighboring nuclei. In live embryos, this segmentation algorithm was used for nc 13 and 14. For nc 11 and 12, nuclei were detected by choosing the center of the nucleus manually. Each of the manually-detected nuclei was then taken to be a disc, 4.4 microns in diameter, centered at this point.

After detection of the nuclei, the Dorsal nuclear gradient was calculated based on previous methods (Liberman et al., 2009). Briefly, the Dorsal gradient concentration in each nucleus was the average intensity of the Dorsal channel for that nucleus divided by the average intensity of the histone channel for that nucleus, multiplied by the mean intensity of all of the nuclei. Nuclear Sna-GFP intensities were calculated in a similar manner.

*Characterizing the Dorsal gradient*

Each measurement of the Dorsal gradient was fit to either Equation 1 (all embryos besides Figure 6) or Equation 2 (embryos depicted in Figure 6), with the $x^2$ term replaced by $(x-\mu)^2$, where $\mu$ is the unknown location of the ventral midline. Matlab's curve-fitting function "fit" was used, using nonlinear least squares and the following starting guesses: for gradient amplitude, the difference between the maximum intensity nucleus and the minimum intensity nucleus; for basal levels, the minimum intensity nucleus; for the

location of the ventral midline, the location of the highest intensity nucleus; for the width of the gradient, 0.15; for the slope of the tail (where applicable), zero. The following lower bounds on parameters were used: for gradient amplitude, one-tenth the difference between the maximum intensity nucleus and the minimum intensity nucleus; for basal levels, zero; for the location of the ventral midline, the location of the highest intensity nucleus minus 30% DV location; for the width of the gradient, 0.05; for the slope of the tail (where applicable), $-10^6$. The following upper bounds on parameters were used: for gradient amplitude, ten times the difference between the maximum intensity nucleus and the minimum intensity nucleus; for basal levels, the average between the maximum intensity nucleus and the minimum intensity nucleus; for the location of the ventral midline, the location of the highest intensity nucleus plus 30% DV location; for the width of the gradient, 1; for the slope of the tail (where applicable), $10^6$. Uncertainties in parameter estimates were taken to be one-half the width of the 68% confidence interval. In particular, the uncertainties in locating the ventral midline of each embryo in Figure 6A were all less than 1% of the DV axis length.

To normalize the Dorsal nuclear gradients in fixed embryos (Figure 6A), the raw Dorsal nuclear gradient for embryo $i$ was subtracted by $B_i$, then was divided by $A_i$, where $A_i$, $B_i$ are the gradient amplitude and basal levels for embryo $i$, respectively. After aligning each of the embryos to their individual ventral midlines (calculated as described above) and normalizing in this fashion, the embryos in Figure 6A were plotted on top of each other.

*Simulations of gradient tail slopes*

The histogram in Figure 6B shows a significant proportion of embryos with positive gradient tail slope. To test whether this is an artifact of our image analysis procedure, in particular of the normalization by the nuclear intensity, the following control simulations were done. The average non-uniformity in the nuclear intensity from our dataset was calculated. The average non-uniformity had a peak value of 162% of the median nuclear intensity, and a minimum value of 72.79% of the median nuclear intensity. This non-uniformity was modeled as a circular normal distribution with a concentration parameter of 8.16 (red curve, Figure S4C). We normalized a Gaussian-like curve, with flat gradient tails (blue curve in Figure S4C), by this non-uniform nuclear intensity, resulting in the green curve in Figure S4C. The green curve was then subjected to the same fitting procedure as our real data, and the normalized slope of the gradient tail was found. This procedure was performed N = 160 times, each with a random placement of the location of the peak in the nuclear intensity (DV position = 0.28 in the example shown in Figure S4C, red curve). This resulted in the histogram of normalized gradient tail slopes found in Figure S4E. The results show that the histogram has a slightly positive bias, but is mostly evenly distributed around zero. This is markedly different from the histogram in Figure 6B, and this control simulation procedure shows the gradient tail slopes calculated from our fixed embryo data are not an artifact of the image analysis procedure.

*Measuring gene expression profiles*

To obtain semi-quantitative data of the location of gene expression (that is, data that contains relative intensities, but not absolute intensities), first the ventral midlines of the

fixed embryos in Figures 4-7 were found manually. Next, each gene expression profile was background-subtracted and normalized for laser power (see below). Then, assuming symmetry about the ventral midline, each gene expression profile was split into two, corresponding to the right and left sides of the embryo. Finally, gene expression profiles for each nuclear cycle or nuclear cycle substage were averaged together.

The locations of the gene expression boundaries, as found in Figure 7I,L,O, were quantified as described previously (Liberman et al., 2009). Briefly, the gene expression pattern for each gene in each embryo was fitted to "canonical" gene expression patterns based on changing the heights, widths, and locations of the canonical patterns. Once best-fit canonical gene expression patterns were found, gene expression boundary locations were defined as the locations where the canonical pattern reached half-maximal intensity.

*Analysis of intronic sog*

For intronic *sog*, the intensity of the nuclear dots (nascent transcripts) was found in the following manner. First, in the intronic *sog* color channel, the max intensity pixel in each nucleus was found. To ensure this pixel was not the effect of a single improbable photon, the median intensity of the 3-by-3 neighborhood centered on this pixel was taken as the intensity of the nuclear dot.

Because the profiles of nascent transcripts are salt-and-pepper (see Figure S3F), this was translated into a smooth profile in the following manner. First, the locations of the nuclei (in normalized DV coordinates) were placed into bins on a mesh from zero to one with 40 points. The value of the non-smoothed profile at bin $i$ was taken as the max intensity seen in a window 5 bins wide, centered at bin $i$. If a bin contained zero nuclei,

this mesh point was given a value corresponding to the average of the two intensities at the nearest two mesh points with at least one nucleus, with the provision that the two mesh points had to be on different sides of the point with zero nuclei. In other words, go left until you find a mesh point with a value, and then right until you find one with a value, and take the average of those two.

This procedure will clearly give you a profile that is too broad, so to narrow it back down to the correct size, we morphologically eroded the profile with a structuring element of width 5 points (to counteract the previous sliding window of width 5 points). After this procedure, the non-smoothed profiles were smoothed using a sliding window of width 5 points (an averaging procedure).

*Background subtraction of gene expression profiles and Sna-GFP profiles*

A precise, quantitative background is difficult to measure on embryos that have been manually cross sectioned. This is because the optical section taken using confocal microscopy must pass through varying thicknesses of physically-damaged tissue. However, a rough estimate of the background intensity of fluorescent in situ hybridization images was determined in the following manner. First, control wildtype embryos were taken through the fluorescent in situ hybridization protocol, but no anti-sense riboprobes were added. However, the primary and secondary antibodies were used consistently. For example, for *sna*, the hapten used with the anti-sense riboprobe was biotin, with anti-biotin raised in goat as primary antibody, and anti-goat (raised in donkey) conjugated with Alexa Fluor 488 was used as secondary antibody. Therefore, the background experiment for *sna* included embryos treated with those two antibodies. n = 8 (for *sog*, *zen*) or n = 16 (for *vnd*,

*sna*) of these embryos were imaged, using the same microscope conditions as were used for experimental embryos, with the exception of changing laser power (see below). Intensity profiles from these images were found using the same image analysis procedure as for the experimental embryos. The average background intensity profiles are plotted in gray in Figure S2A, with errorbars representing the standard deviation of all background intensity profiles for that gene.

Once background intensity profiles were found, they were applied to the experimental data in the following manner. The "structural background" of each gene expression intensity profile was found through a morphological opening using an appropriately large structuring element (for Type III genes, 60% of the embryo perimeter; for others, 40% of the embryo perimeter). This structural background can be thought of as the intensity of the profile outside of its normally-accepted expression domain. For example, with *sna*, this would roughly be the intensity of the profile from DV position = 0.20 to 1. If this structural background was statistically greater than the background intensity from the control embryos, then the background from the control embryos was used. If not, then the structural background was used (in which cases the structural background is likely to correspond to a true lack of gene expression). This is because of the uncertainty in comparing embryo-to-embryo when sectioning manually can sometimes lead to structural backgrounds less than the background from control embryos.

The background-subtracted (and normalized; see below) profiles were plotted in Figure 4F, and were also used in the fitting procedure for the mRNA dynamics model (see below).

For Sna-GFP, a similar procedure was used, with the control embryos being wildtype (that lack Sna-GFP).

*Correction for laser power*

To correct for embryos imaged on different days, laser power baselines for each day were taken during each imaging session. Using those data, gene expression profiles, as well as Sna-GFP profiles, were normalized according to the laser power used to image them (Liberman et al., 2009). This also allowed us to capture a full dynamic range for embryos with drastically different intensities (such as control embryos or early embryos that contained no gene expression vs. embryos displaying bright, "mature" gene expression profiles).

*Normalization of gene expression profiles*

In addition to being background-subtracted and laser power corrected, the gene expression profiles of *sna*, s*og*, *vnd*, and *zen* shown in Figure 4F (and also used for data-fitting; see below) were normalized such that the peak intensity was equal to one. This was done in the following manner. After average profiles were found for each gene and each nuclear cycle substage, they were background subtracted. Then, for each gene, the peak intensity across all nuclear cycle substages was set to one. For example, for *sna*, the peak intensity for all nuclear cycle substages occurred during late nc 14 at roughly $x = 0.05$ (see Figure 4F). All of the averaged *sna* profiles (across all nuclear cycle substages) were then divided by this intensity.

*Averaging of the three live Dorsal-Venus nuclear gradient time series*

The three live Dorsal-Venus nuclear gradient time series (Figure S1B-D) were averaged together in the following manner. Each nuclear cycle interphase was treated independently, as was each nuclear cycle mitosis. As an example, the length of nc 11 interphase from each embryo was determined based on the "saw-tooth" pattern of the gradient amplitude. Due to small variations in development time, these lengths were slightly different. The gradient amplitudes, basal levels, and widths during nc 11 interphase were plotted together after stretching or shrinking the duration of nc 11 interphase of each individual embryo to fit the average duration of nc 11 interphase (Figure S1F-H). This same stretching/shrinking/averaging procedure was performed on each interphase and mitosis. Afterwards, the gradient amplitudes, basal levels, and the gradient widths were averaged together to arrive at an averaged Dorsal-Venus nuclear gradient (black curves in Figure S1F-H). The background levels (gray curve with errorbars in Figure S1B-D,G) were measured from embryos carrying only H2A-RFP (and not Dorsal-Venus) that were imaged in the same manner as the embryos in Figure S1B-D.

*Simulation of the Dorsal gradient*

An estimate of the wildtype Dorsal gradient was constructed from live imaging time series data of Dorsal-Venus and nc 14 fixed tissue data in the following manner. The averaged gradient amplitude and basal levels from live Dorsal-Venus data (see above) were used as $A(t)$ and $B(t)$. Due to the fact that the Dorsal-Venus nuclear gradient is measurably wider than the wildtype Dorsal nuclear gradient (Figure 1D-F), and that live imaging showed the gradient width to be constant in time, the width was taken to be 0.14, the mean

of the nc 14 fixed tissue data set (Figures 1F and 6A). The slope of the gradient tail was assumed to be equal to $M(t) = -0.1A(t)$ to reflect the mean value of the normalized slope of the gradient tail. After $A(t)$, $B(t)$, σ, and $M(t)$ were computed, the simulated wildtype Dorsal nuclear gradient was computed according to Equation 2, with 10% Gaussian noise added to each point in space and time. For example, the noise added to the gradient $c(x,t)$ at $x = x_0$ and $t = t_0$ was randomly chosen from a normal distribution with mean $c(x_0,t_0)$ and standard deviation $c(x_0,t_0)/10$. The rationale for including noise in the Dorsal gradient is because without relevant biological noise, a deterministic model can read an arbitrarily shallow slope with perfect precision. A level of 10% noise was chosen as suggested by previous studies of morphogen gradient precision (Gregor et al., 2007a). For an exploration of the effect of this noise on gene expression patterns, see Figure S5.

*Statistical analyses*

The widths of some pairings of the populations of embryos in Figure 1F were shown to have statistically different means by either modified *t*-test (Welch, 1947) for differences between wildtype cross sections and anti-Dorsal stainings of *dl1/+; dl-venus/+*, or by *t*-test for correlated samples for differences between anti-Dorsal and anti-GFP within the same embryos.

*Time delays of mRNA production*

The four genes analyzed here (*sna*, *vnd*, *sog*, *zen*) have gene lengths of 1676, 6780, 21970, and 1330 bp, respectively. At a transcription rate of 1.1 kb/min (Thummel et al., 1990), this would mean time delays of 1.52, 6.16, 19.97, and 1.21 minutes, respectively.

At nearly 20 minutes for *sog*, we would not expect *sog* expression before mid nc 14, as unfinished transcripts are not "carried over" through mitosis into the next nc interphase, but are instead degraded (Rothe et al., 1992; Shermoen and O'Farrell, 1991). Even the fastest reported transcription rate for the early embryo (1.4 kb/min, Shermoen and O'Farrell, 1991) would result in a delay of almost 16 minutes for *sog*, which again is prohibitively long for the appearance of mature *sog* transcripts before mid nc 14. Since *sog* transcripts were seen in most of the embryos from late nc 13, for our model we assumed the transcription rate for all genes was fast enough such that *sog* would be just completed half way through nc 13 interphase (6.15 min, corresponding to 3.5 kb/min). While this may be a questionable assumption, it is consistent with the earlier than otherwise expected appearance of *sog* transcript.

*Staging of fixed embryos*

Nuclear cycle 13 was divided equally into early and late substages. nc 14 lasts around 45 minutes at 25ºC and was thus separated into three different substages based on nuclear morphology: early, mid, and late. Comparing the nuclear morphology of fixed, cross-section embryos with that of H2A-RFP in live embryos, it was determined that the binning procedure resulted in the early and late nc 14 stages being ~10 minutes in duration or ~20% of nc 14. Mid nc 14 was the longest at ~30 minutes or ~60% of nc 14.

*Fitting model parameters to gene expression data*

The semi-automated fitting procedure took place as follows. First, for *sna*, the model parameters $\theta_{sna}$ (the Dorsal signaling threshold to activate *sna* gene expression) and

$\tau_{sna}$ (the lifetime of *sna* gene product) were optimized against gene expression data for *sna*, as depicted in Figure 4F.

The initial guess for the value of the threshold $\theta_{sna}$ corresponded to the value of the simulated Dorsal levels at $x$ = 0.20 at a timepoint when the gradient amplitude was maximal in nc 14. Allowing for 20% error in this presumptive gene expression boundary (that is, $x$ between 0.16 and 0.24), upper and lower bounds on the possible value of $\theta_{sna}$ were chosen. Because the threshold corresponded to a binary switch, a fitting procedure for this parameter using Newton's method on the gradient of the objective function was inherently unstable. Therefore, thirty values of the threshold were chosen (with a uniform distribution between the upper and lower bounds), and fifty values of mRNA lifetime were chosen (uniformly-distributed on a log scale between 1 and 1000 minutes).

The objective function $\gamma$ was a chi-square function:

$$\gamma = \sum_{i=1}^{N} \frac{(Y_i - \hat{y}_i)^2}{\sigma_i^2},$$

where N=151 was the number of points along the discretized DV coordinate; $Y_i$ was the background-subtracted, normalized average value of the measured gene expression profile at $x_i$; $\hat{y}_i$ was the value of the simulated gene expression profile for the current choice of $\theta_{sna}$, $\tau_{sna}$ at $x_i$; and $\sigma_i$ was the standard error of the mean for the measured gene expression profiles at $x_i$.

For each choice of fixed ($\theta_{sna}$, $\tau_{sna}$), a value $\gamma$ was calculated. The choice of ($\theta_{sna}$ , $\tau_{sna}$) that resulted in the smallest value of $\gamma$ was taken as the best-fit parameters for *sna*.

Once the optimum values for $\theta_{sna}$ and $\tau_{sna}$ were found, the same optimization procedure was performed to determine ($\theta_{vnd}$, $\tau_{vnd}$) and ($\theta_{sog}$, $\tau_{sog}$) independently. However,

the optimization procedure was unstable for these genes in that the program did not converge on the best-fit mRNA lifetime. Therefore, the parameters were manually adjusted "by-eye" such that simulated gene expression matched the *vnd* and *sog* patterns in late nc 13 through late nc 14.

The same optimization procedure was performed on the equation for *zen*, with the exception that only observed data from late nc 13 and early nc 14 were used to fit the parameters, because by mid nc 14, the Dpp-dependent expression of *zen* was already quite pronounced.

The final outputs of the fitting procedure showed mRNA lifetimes to be 12 minutes for *sna*, 10 minutes for *sog*, 10 minutes for *vnd*, and 12 minutes for *zen*.

*zen repression of sog*

For Figure S5, a case was considered in which a dorsally-acting factor repressed *sog*. This was motivated by the fact that, if the *sog* threshold was too low, *sog* was expressed strongly in all nuclei (except those in which it is repressed by *sna*), and if the threshold was made only slightly higher, the domain of *sog* expression became too narrow. This implies that there may be a dorsally-acting factor that represses *sog*. In this case, we allowed *zen* to repress *sog* strongly in the same manner that *sna* repressed *sog* ventrally, with $K_{zen} = 0.05$.

APPENDICES

# B. Supplementary Materials for Chapter III

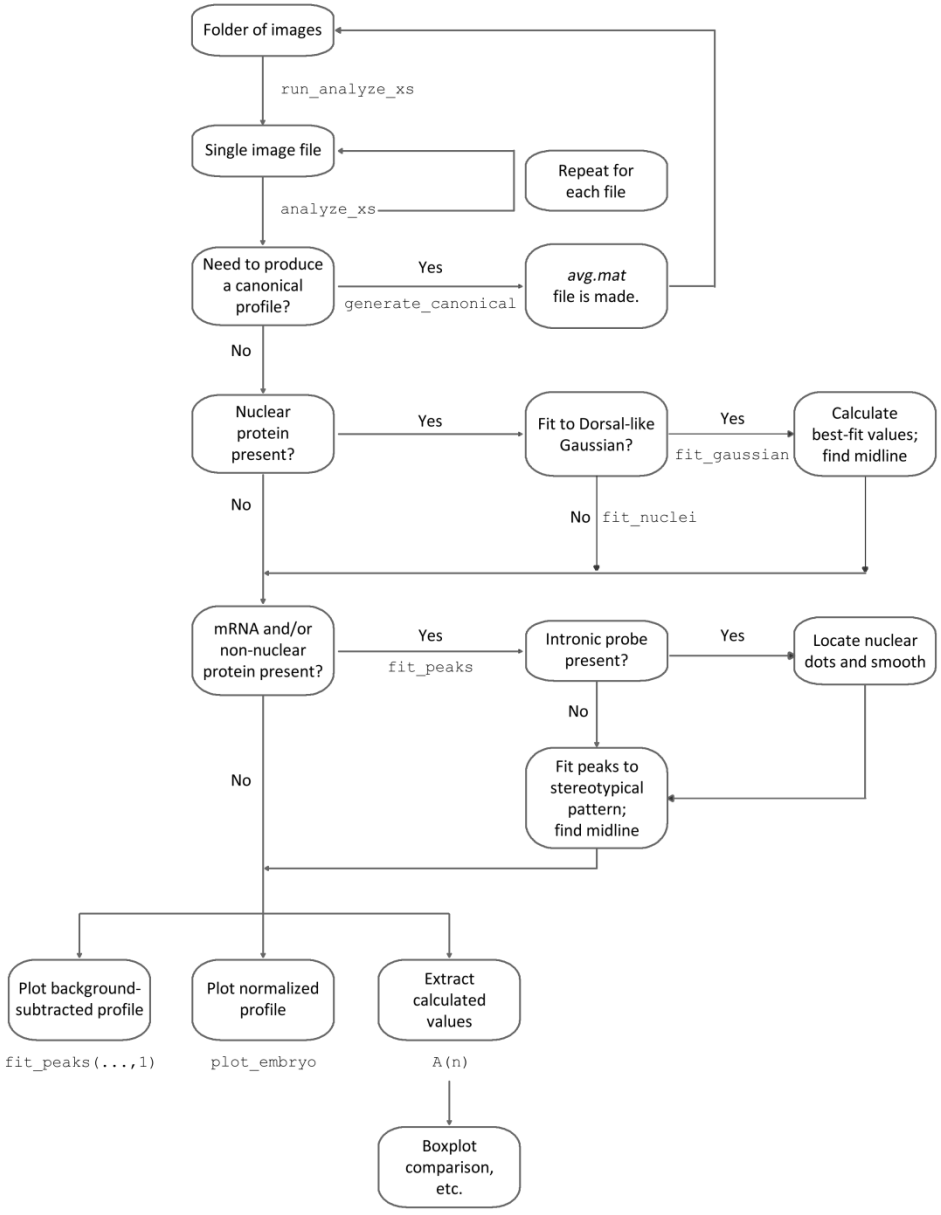
# SUPPLEMENTARY INVENTORY

# S1. User's manual for the package of image analysis and data fitting code

## S1.1. OVERVIEW

`run_analyze_xs` is used to extract data from protein/mRNA stainings of *Drosophila* embryos and characterize their expression. The input is a folder of fluorescent images of cross-sectioned (xs) embryos. It returns a structure array of quantified values for each staining(s) in each embryo. Accompanying functions display the information in graphical form.

## S1.2. WORKFLOW



## S1.3. SYNTAX

```
A = run_analyze_xs( 'folder', [channels], {'channelnames'}, 'genotype')
```

The returned value `A` is an *nx1* structure array with *n* corresponding to the number of files in the folder that were completed by the analysis.

## S1.4. TYPES OF FILES

This function accepts LSM files, with each LSM file containing a single 3D cross-sectioned embryo. Imaging settings are generally 1.3 μm per slice, 15 slices per z-stack (see section 2.3 for additional details). It is imperative that all the files in the folder being analyzed contain identical staining type in identical channel order (see below).

## S1.5. SIGNIFICANCE OF INPUT VARIABLES

The input parameter `channels` must be a numeric row vector with the same number of elements as there are channels in the image. See table below for description of the types of channels and their numeric codes.

`channelnames` must be a single row cell array variable with the same number of elements as there are channels in the image. Each element of the cell array must be a single row string that names the molecular species detected in that channel, such as `'sna'`. If more than one molecular species is in a given channel, separate them by commas, with no spaces, such as `'ind,sna'`. The names of the molecular species must be a member of a given list of known names. This is in particular important for the `fit_peaks` function, which operates on mRNA and intronic channels. See Sections S1.5.2, S1.6, and S2.9 for more details.

The input genotype must be a single row string.

The matching of `channels` and `channelnames` is critical for the program to run properly. Options for input that are made available in this package are listed in the table below.

| channels | DESCRIPTION | Channelnames |
|---|---|---|
| 1 | nuclei | H3, DAPI, etc. |
| 2 | non-nuclear protein or mRNA | sna, vnd, sog, zen, etc. |
| 3 | nuclear protein | dl |
| 4 | intronic probe | sna, vnd, sog, zen, etc. |
| 5 | none of the above | Brightfield, N/A |

*S1.5.1. nuclei*

Nuclear staining, such as with histone H3 antibody, is required for analysis of nuclear proteins and nascent transcripts (intronic probe).

*S1.5.2. non-nuclear protein or mRNA*

This analysis requires a prior estimate of the gene/protein canonical expression. See `Geneaverages` folder for complete list of currently existing genes and their proper call in format `<name>avg.mat`. New genes require `avg` file to be made (unless a similar `avg` profile can be substituted) in order to run `fit_peaks`. The supplement folder *Functions>Geneaverages* lists the profiles included. See S1.8.8 on generating a new canonical profile.

*S1.5.3. nuclear protein*

Nuclear proteins labeled `dl` are assumed to have a Gaussian distribution and are analyzed accordingly. Otherwise, nuclear proteins are fit using a canonical profile via `fit_nuclei`.

*S1.5.4. intronic probe*

Nascent transcripts can be detected using an intronic probe. They appear as nuclear dots, and `analyze_xs` locates maximum intensity pixels within nuclei and measures the strength of the nuclear dot. `fit_peaks` fits a smoothed version of the data to estimated peaks and therefore also requires an `avg.mat` file to exist (see section S1.5.2).

*S1.5.5. none of the above*

Any image channel that does not contain any of the above molecular species. This includes brightfield images.

## S1.6. TIPS

Several obstacles commonly encountered when running the program, and how to avoid them, are listed below.

Varying from `channelnames` (ex. `'snail'` instead of `'sna'`) will cause errors. See Section S1.5.2.

- It is important that the input is in the same sequence as the saved file. Check the order in which channels, and thus the corresponding stainings, are saved. The sequence may be independent of channel wavelength. One method is to load the LSM file onto ImageJ.
- Bright spots near the embryo (ex. pieces of dust) will cause the program to abort.
- Nuclear cycles 11 and 12 of the Drosophila syncytium require manual input for nuclear segmentation. This is due to the fact that nuclei are sparse at this stage.

Please note that if running `fit_peaks(…,1)`:

- The function requires an existing folder named "*Fittedpeaksimages*" in the Methods folder. Output JPEGs will be saved here.
- Nuclear proteins are not plotted.
- Currently, the function cannot create the "*Fittedpeaksimages*" folder; it must be a preexisting folder.
- See section S1.8.4 for more information.

If `fit_gaussian` or `fit_peaks` returns errors, user should be able to manually skip to `plot_embryo`. However:

- Calculated values will have missing fields
- Profiles will not be centered such that the ventral midline is at 0. User can still manually choose the midline. See section S1.8.7.

## S1.7. UNDERSTANDING THE ANALYSIS FUNCTIONS

Several functions are called throughout the course of this analysis, and they are listed in the tables below.

*S1.7.1. From `run_analyze_xs`:*

| FUNCTION | DESCRIPTION |
|---|---|
| `analyze_xs` | Obtains the intensities of the image channels of interest. |
| `fit_gaussian` | Fits the nuclear protein distribution (in particular the Dorsal nuclear gradient data) to a Gaussian to fine the best-fit values of the gradient amplitude, basal levels, width, and midline. |
| `fit_peaks` | Fits gene expression pattern(s) to stereotypical peaks of gene expression. This is probably the most complicated of the analysis functions. Uses `canonicalgeneborders`, `circshiftDU`, `find_midline`, `smooth_intron`, `subtrbkgrnd`, `cell2str`, `str2cell`, `isodd`, `strfindDU`, `genefit`, `num2strDU`. |

*S1.7.2. From `analyze_xs`:*

| FUNCTION | DESCRIPTION |
|---|---|
| `lsminfo` | Reads meta data about the LSM file. |
| `lsmRead2` | Loads z-stack images from the LSM file. |

| | |
|---|---|
| `borderFinder` | Locates the periphery of the embryo in each z-slice. Uses `c1628`, `circfit`, `conseccheck`, `gaussfiltDU`, `roundx`. |
| `domainMeas` | Measures the intensity of the color channels in quadrilaterals placed around the periphery of the embryo. This is for all color channels but in particular it is important for mRNA. This is always called by default, even if the user does not specify mRNA as one of the color channels. Uses `roundx`. |
| `find_nuclei` | Locates nuclei. This is only called if the user specifies one of the channels as nuclei. Uses `unroll2` (uses periphery points to unroll nuclear layer into a strip). |
| `nuclearintensity` | Finds the intensity of the nuclear proteins in the specified color channels, just on top of the nuclei (that was found by `find_nuclei`). This is only called if the user specifies one of the color channels as nuclei and at least one of the color channels as a nuclear protein. |
| `intronicintensity` | Finds the max intensity dot on each nucleus, which corresponds to the intensity of the intronic probe (i.e., nascent transcript or "nuclear dot"). This is only called if the user specifies one of the color channels as nuclei and at least one of the color channels as an intronic probe. |

*S1.7.3. From `fit_peaks`:*

| FUNCTION | DESCRIPTION |
|---|---|
| `canonicalgeneborders` | Loads a "canonical" version of the gene expression peak of the gene-of-interest and calculates where the borders of that gene expression peak are, based on an h-maximal threshold (usually h=1/2). |
| `find_midline` | Finds the most likely location of the midline based on the patterns of gene expression as well as what types of genes are in each channel. |
| `smooth_intron` | Takes the salt-and-pepper pattern of intronic probe intensity and smooths it out so the profile can be easily fitted to a canonical gene expression peak. |
| `genefit` | Loads in a canonical gene expression peak for the gene provided and returns what the real gene expression peak looks like based on changing amplitude, background level, location of peak, and a stretching factor. |

*S1.7.4. Plotting the data*

Upon the completion of `run_analysis_xs`, user may run the function `plot_embryo` described below.

| | |
|---|---|
| `plot_embryo` | Plots all of the data extracted by `analyze_xs`. Uses the automagically-detected midline from one of the "fit" functions to center the plot. If the midline has not been previously found, the user can interact with the plot and choose a midline (see section S1.8.7). Uses `circshiftDU`, `mDU`. |

## S1.8. EXAMPLES

Below are several examples of inputs into the Matlab command window and corresponding outputs.

*S1.8.1. Example 1A – Going through `run_analyze_xs`*

For a folder named *embryos.mdb* with five LSM files of five embryos stained with *vnd* gene in the first channel, Dorsal nuclear protein in the second channel, and histone H3 nuclear protein in the third channel:

**Input 1A:**

```
>> X = run_analyze_xs('embryos.mdb', [2 3 1], {'vnd' 'dl' 'H3'}, 'wt');
```

**Output 1A:**

```
j = 1
j = 2
j = 3
e = 1
j = 4
```

The return value `X` is a 4x1 structure array. The 4th embryo in the folder gave error(s) and returned `e=1` which is not included in `X`. The error can be determined by loading a mat file that is automatically saved by `run_analyze_xs`. The mat file is named `run_analyze_xsError.mat` by default, but can take on an appended portion of the name if "`outfile`" is specified in `run_analyze_xs` as the 5th argument. Also, results of the current analysis are saved to a mat file called `run_analyze_xsCurrent.mat`. The same appended naming can be performed here. This file is updated after each successful embryo analysis.

*S1.8.2. Example 1B*

To examine the data of an individual embryo `n`, enter `X(n)`. The input below calls on the second embryo.

**Input 1B:**

```
>> X(2)
```

**Output 1B:**

The output is grouped into four different categories. The first is the metadata, which is composed of the fields from "`filename`" to "`L_bar`". The second is the secondary data, which are in the fields from "`s_mid`" to "`Intron`". The last two categories are the tertiary data regarding nuclear proteins and that regarding genes. See S2.2.3 for detailed description of fields.

```
ans =
            filename: 'embryo2.lsm'
            channels: [2 3 1]
        channeltypes: {1x3 cell}
        channelnames: {'vnd'  'dl'  'H3'}
            genotype: 'wt'
            metadata: [1x1 struct]
                   H: 742
                   W: 777
                   D: 20
                Lbar: 276.7770
               s_mid: [-0.6933 -0.7046 NaN]
                   s: [301x1 double]
                   t: [301x3 double]
                   S: [1645x1 double]
                   R: [1645x1 double]
               Std_R: [1645x1 double]
              Intron: NaN
    nucprotein_names: {'dl'}            gene_names: {'vnd'  'dl'}
                   A: 1.5784e+03                 sV: [0.1866 0]
                   B: 244.3316                   sD: [0.3142 1.0122]
                   M: 15.8470                     w: [0.1276 1.0122]
                  mu: -0.7046                   dsV: [0.0017 0]
                 sig: 0.1574                    dsD: [0.0018 0.0073]
                  dA: 5.5629                     dw: [0.0035 0.0073]
                  dB: 4.2917               gof_gene: [0.9752 0.9977]
                  dM: 5.8413
                 dmu: 4.3774e-04
                dsig: 5.5205e-04
                 gof: 0.9893
```

See S1.8.8 regarding the presence of the `'dl'` gene.

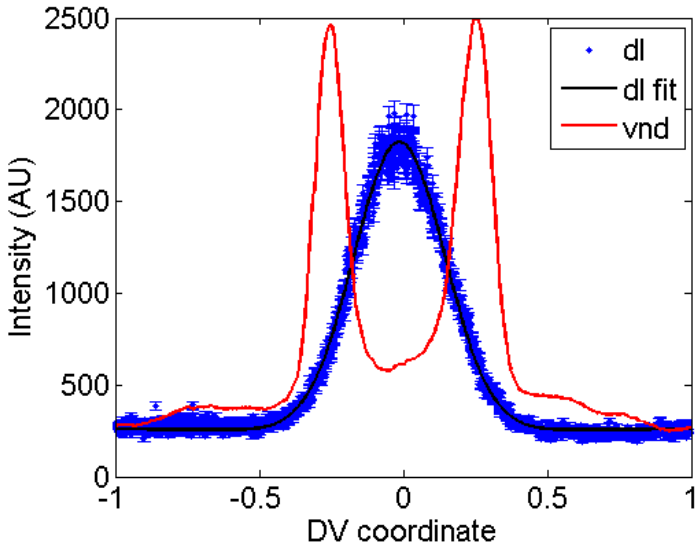
*S1.8.3. Example 1C*

The input here will plot the data of the second embryo using function `plot_embryo`.

**Input 1C:**

```
>> plot_embryo(X(2));
```

**Output 1C:**




*S1.8.4. Example 1D*

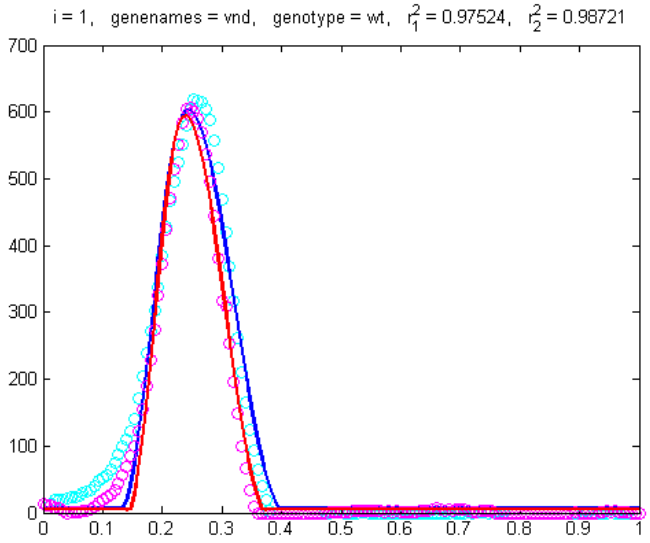
To plot the non-nuclear protein(s) and/or gene(s) using function `fit_peaks`, include the parameter "`1`".

**Input 1D:**

```
>> fit_peaks(X(2),1);
```

**Output 1D:**

JPEG(s) are saved in folder "*Fittedpeaksimages.*" (This folder must be created by the user first; Matlab will return an error if this folder is not found in the working Matlab directory.) Note that Dl is not plotted since it is a nuclear protein and not analyzed by `fit_peaks` (it goes through `fit_gaussian` instead). For gene *vnd*:




These profiles represent half (0 to 1) of the embryo along the DV axis and:

(a) Centered at midline
(b) Background subtracted
(c) Four curves, two open-circles (data) and their corresponding two lines (fitted)
(d) Each pair of curves represents each half of the embryo (-1,0) and (0,+1)

*S1.8.5. Example 1E*

To call on data for all embryos in folder, do not include *n*.

**Input 1E:**

```
>> fit_peaks(X,1);
>> plot_embryo(X);
```

**Output 1E:**

JPEGs for all non-nuclear proteins and genes of all embryos in `X` will be saved in "*Fittedpeaksimages*" folder.

Plots for all embryos in `X` will appear briefly as they are saved to JPEG format.

*S1.8.6. Example 2 – Analyzing a single image file*

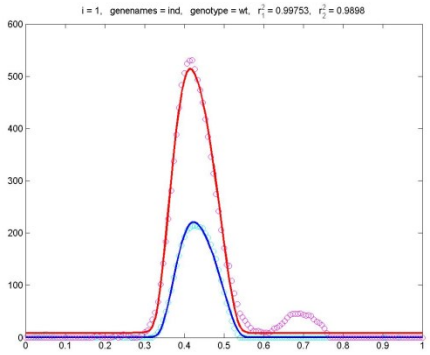
To analyze a single LSM file, user may run the function `analyze_xs` with the same input variables as `run_analyze_xs`. Then, depending on stainings, subsequently run `fit_gaussian` and/or `fit_peaks`. The file *sample1.lsm* (which is included in the Supplementary material) has *sna* in the first channel, *ind* in the second channel, and nuclear H3 in the third channel.

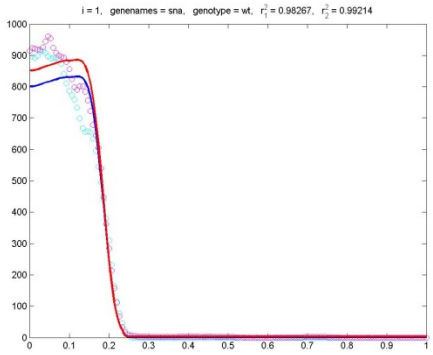
**Input 2:**

```
>> Y = analyze_xs( 'sample1.lsm' , [2 2 1], {'sna' 'ind' 'H3'}, 'wt');
>> Y = fit_peaks(Y);
>> fit_peaks(Y,1);
>> plot_embryo(Y);
```

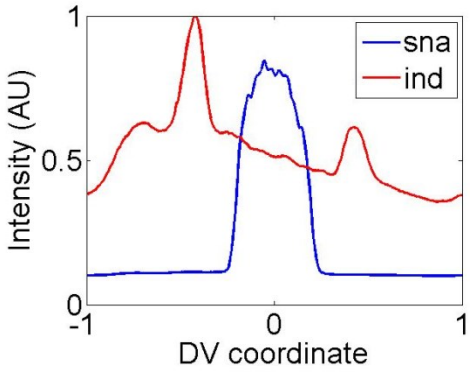
**Output 2:**

There are no nuclear proteins, so `fit_gaussian` was not needed. Files `ind001.jpg` and `sna001.jpg` are saved in folder "`Fittedpeaksimages`":




Centered plot of `ind` and `sna` will appear:



*S1.8.7. Example 3 – Manual centering*

If skipped fitting (`fit_gaussian` and/or `fit_peaks`), or errors returned, user may continue to `plot_embryo` and manually center the profiles. Here, single embryo data `Z` did not go through `fit_peaks`.

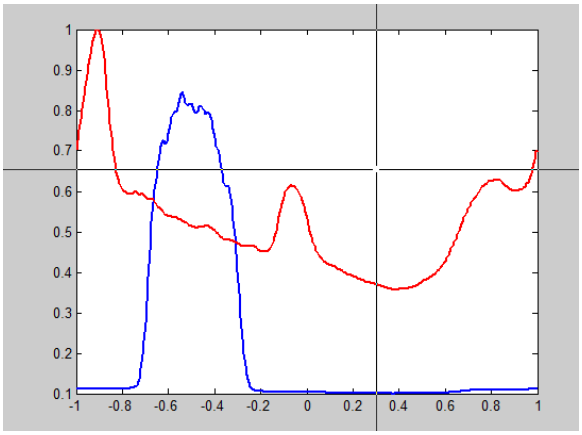
**Input 3:**

```
>> Z = analyze_xs( 'sample1.lsm' , [2 2 1], {'sna' 'ind' 'H3'}, 'wt');
>> [h,L,Z] = plot_embryo(Z);
```

**Output 3:**

Mouse pointer becomes a crosshair: left-click to manually choose where center should be (artificially ~0.3 in example below). If the first choice in clicking on the midline was not quite right, the user can continue to update with more left-clicks. After the user is satisfied in choosing a midline manually, right-click to confirm.

The output variables `h` and `L` refer to the graphics handles (identifiers of each of the curves in the plot) and names of the curves (which go into the legend), respectively. The third output, `Z`, is the original structure with the value of the user-chosen midline saved to the field `s_mid`.



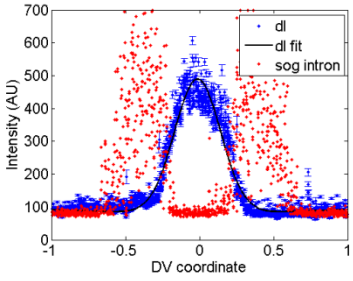
### *S1.8.8. Example 4 – Analyzing nuclear proteins*

Consider a nuclear protein called `abc`.

If `abc` is to be analyzed similarly to Dorsal (i.e., Gaussian-like Equation 1), user should input '`dl`' in `channelnames` since `fit_gaussian` only recognizes '`dl`'.

```
X = analyze_xs('file', [3], {'dl'})
```

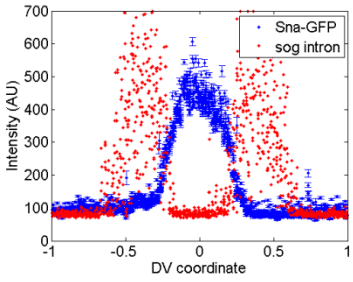
In this case, `fit_gaussian` will calculate the Gaussian properties for `abc`. And `plot_embryo` will display these values as dots with errorbars along with a fitted curve. In addition, `fit_peaks` will also analyze this channel using the *dlavg* canonical profile though it will not appear when calling `plot_embryo`.




However, if `abc` does not fit a Gaussian, it can still be analyzed using `fit_peaks`. Thus it will require an existing canonical profile.

```
X = analyze_xs('file', [3], {'abc'})
```

Even though there is a nuclear protein present as indicated by `[3]`, `fit_gaussian` does not recognize '`abc`' and will not calculate any Gaussian properties (no fitted curve). Instead, `fit_peaks` will use *abcavg.mat* to analyze it.




An image file, *sample2.lsm*, has been provided as part of the Supplementary material and contains a nuclear protein that can be analyzed as '`dl`' or '`Sna-GFP`' (with *Sna-GFPavg.mat*).

### *S1.8.9. Example 5 – Generating a canonical profile*

When analyzing a species for the first time, a canonical profile (<*gene*>*avg.mat*) must be generated to fit the data using `fit_peaks`. Producing an additional canonical plot is also useful when there are changes in the expression domain, for example due to temporal dynamics and/or mutant phenotypes. Nuclear proteins that do not necessarily fit a Dorsal Gaussian distribution can also be fitted to a canonical plot (instead of running the data through `fit_gaussian`). However, these canonical plots do not contain the parameters used to describe Gaussian curves such as amplitude and sigma.

The more samples that are taken into account, the better the canonical profile will be representative. The suggested minimum is 10 images. Multiple profiles may be required, depending on the unique patterns that could change over time, probe location, etc.

**Input 5A:**

```
>> [h,L,soln] = plot_embryo(X);
```

The input `X` is the structure returned from the function `analyze_xs`.

Where:

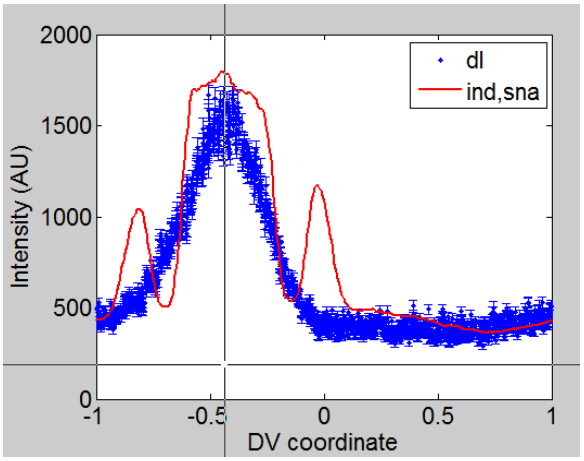
- `h` is a cell array of embryos in `X`. Each element corresponds to all of the object handles of respective plots.
- `L` is a cell array of embryos containing the names of each object in each plot, e.g., a legend can be generated on `i`'th plot by issuing

  ```
  >> legend(L{i})
  ```

- `soln` is the updated structure `X`. `X` and `soln` will only differ when midline was found by hand. In this case, midline values are stored in every channel of `s_mid`.

**Output 5A:**

Similar to Example 3 (0), a window will appear with a data plot of one embryo and the user must manually choose the ventral midline (right mouse click). This is repeated for all embryos in the structure `X`.




Here, the blue curve is Dorsal nuclear protein. Red is `sna` and `ind`, with non-overlapping domains, stained in the same channel. The mouse's crosshair indicates the chosen ventral midline at around -0.3. This midline is saved in the output `soln` and used in the next step with `generate_canonical`.

**Input 5B:**

```
>> [s,t,sD,sV,s_offset] = generate_canonical(soln,'sna','ventral', rbr) ;
```

Where:

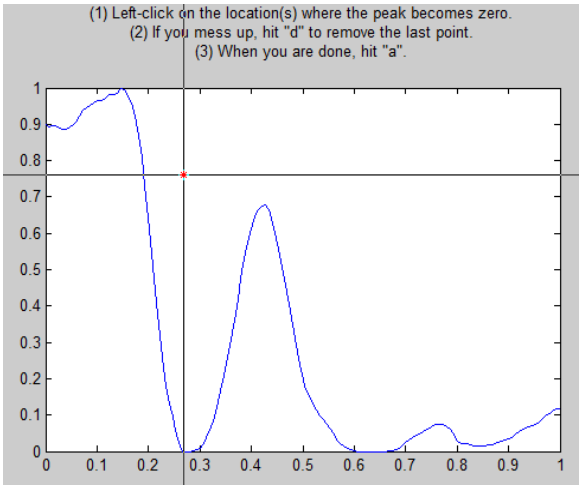
- `soln` is the output of `plot_embryo` in Example 4A
- `sna` is the name of the gene you are making the canonical profile for. For each embryo, the program will find which color channel has `sna` in it, and use that as part of the averaged profile that makes up the canonical profile. This will also be used to name the `<gene>avg.mat` file, i.e., `snaavg.mat`. Note that once this file is made, and `fit_peaks` is later called to fit your gene expression data to a canonical profile, every measured gene that is labeled as '`sna`' will be fit to this profile.
- `ventral` indicates the gene type. With '`ventral`' only one border will be calculated. Other options include '`lateral`' (two borders) and '`dorsal`' (one border). In the case this is not a DV cross section (this could include, for example, the animal-vegetal axis in *Xenopus* embryos, the anterior-posterior axis in saggital sections of *Drosophila* embryos, or the oral-aboral axis of sea urchin embryos), '`ventral`' would refer to a gene whose expression domain includes the point chosen for '`s_mid`'. If you align an anterior-posterior saggital section of a *Drosophila* embryo with the anterior pole at `x = 0`, then hunchback would qualify as a '`ventral`' gene. On the other hand, '`dorsal`' would then refer to a gene whose expression domain includes the point opposite what is chosen for `s_mid`. '`lateral`' refers to a gene whose expression domain includes neither the point chosen for '`s_mid`', nor the point opposite.

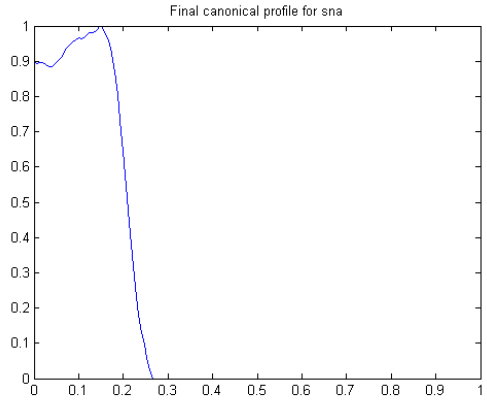
- `rbr` is the radius of the structuring element with which the background is subtracted. Choose this value to be slightly wider than your peak, in units of relative axis length.
- Optionally, not shown, a midline channel index can be provided as the last input. Each channel can have an estimate of the midline. This index indicates which channel to use as the estimate for the ventral midline. The default value is `1`. However, if the midline estimate from channel `1` contains `NaN` (a special Matlab name for "not-a-number"), the program will look for the first channel that contains a non-`NaN` value for the midline.

**Output 5B:**

Since both `sna` and `ind` genes are on the same channel, both curves appear in the plot. However, the inputs '`sna`' and '`ventral`' indicate that only one border should be chosen, and that border should be the border of `sna`. Thus, the user can only specify one point where the peak becomes zero. This is shown in the left plot where the zero point is taken at `~0.25`.

The zero point is chosen by clicking on the left button of the mouse and a red dot will appear. Once all the point(s) are selected, enter '`a`' to finish. If you would like select a different point, enter '`d`' to remove the previous point and try again.




Everything dorsal to that point is set to zero, and this final canonical profile is saved in the working folder (plot on right). In this example, the file is named `snaavg.mat`. If you have a folder where all of your canonical profiles are stored (i.e., *Functions>Geneaverages*), then move `snaavg.mat` to that folder.

The output variables are:

- `s` is the domain values (x-axis)
- `t` holds the values of the curve (y-axis)
- `sD` is the location of the dorsal border (or the border distal to the location chosen as s_mid) of the canonical peak. If the gene is a `dorsal` gene, then `sD = 1`, by definition.
- `sV` is the location of the ventral border (or the border proximal to the location chosen as `s_mid`) of the canonical peak. If the gene is a '`ventral`' gene, then `sV = 0`, by definition. `s_peak` is the maximum value for '`lateral`'. Note that for '`dorsal`' `s_peak = 1` and for '`ventral`' `s_peak = 0`.

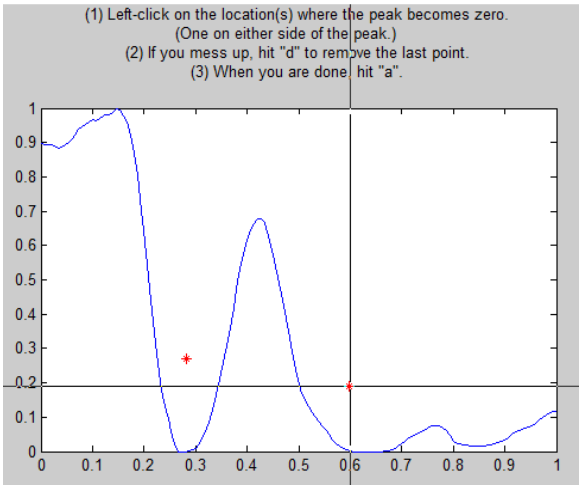
These variables are all stored in the file `<gene>avg.mat`. In addition to these variables, `rbr` and '`s_offset`' are also stored in `<gene>avg.mat`. `rbr` is described above. `s_offset` is the location of the peak of the canonical.

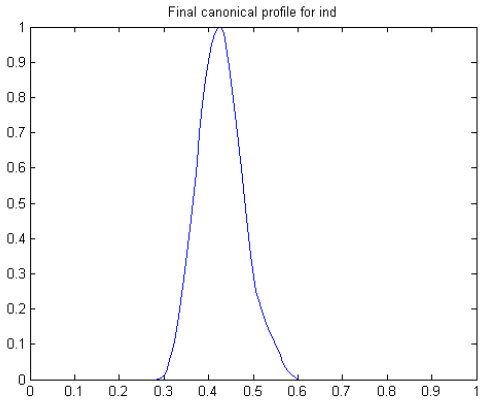
**Input 5C:**

```
>> [s,t,sD,sV,s_offset] = generate_canonical(soln,'ind','lateral', 0.25);
```

**Output 5C:**

For a lateral gene, two borders are required: the ventral border and the dorsal border. Thus, two points are required from the user. Everything outside the region in between these points is set to zero.




In addition, the profile is normalized such that the peak is at 1.

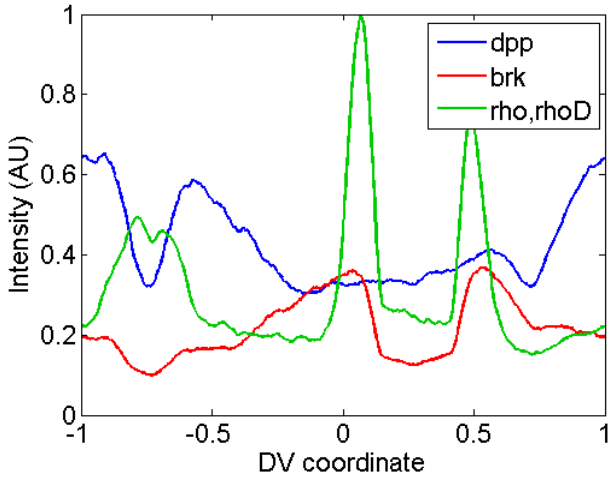
*S1.8.10. Example 6 – Analyzing multiple genes in a single channel*

If multiple genes with non-overlapping expression domains are put in a single channel, this program is able to analyze each expression domain separately. To do this, you need to tell `analyze_xs` that multiple genes are present in the same channel by separating them with a comma. For example, if `sna` and `ind` are in the same color channel, then the `channelname` would be '`ind,sna`'. This feature can be extended to genes that have multiple domains of expression, such as the gap genes, or `rho`. In the following example, we demonstrate how to run the analysis on `rho`, which has a lateral domain ('`rho`'), and a dorsal domain ('`rhoD`'). Note that these two domains are identified separately and require unique canonical profiles.

**Input 6A:**

```
>> soln = run_analyze_xs('embryos2.mdb',[2 2 2],{'dpp','brk','rho,rhoD'},
'wt');
>> plot_embryo(soln(8));
```

**Output 6A:**



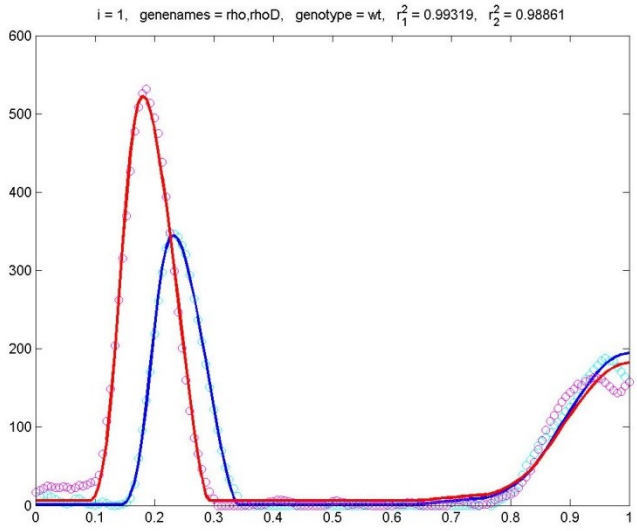

The ventral midline has not yet been established, but there are two distinct rho domains: (1) lateral expression around 0 and 0.5 and (2) dorsal expression around -0.75.

**Input 6B:**

```
>> fit_peaks(soln(8), 1)
```

**Output 6B:**

The lateral domains of rho are detected by the program at approximately 0.1-0.3 DV position, and the dorsal domains are detected correctly as well.



*S1.8.11. Example 7 – Generating a boxplot*

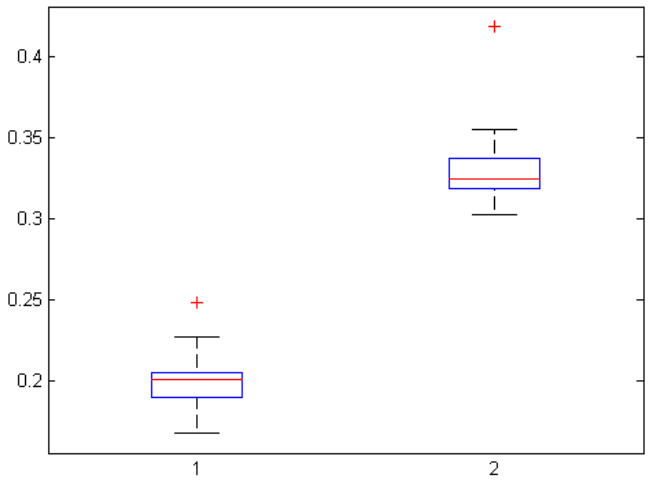
Once analysis of the images is complete, the numbers assigned can be useful in comparing different gene boundaries, identifying mutants, etc. One method is using a boxplot to visualize the location of domain boundaries.

Here, we plot the ventral and dorsal boundaries for the gene *vnd* in wildtype embryos. Again, the embryo is measured such that 0 is the ventral-most point and 1 is the dorsal-most point.

**Input 7:**

```
>> v = [soln(:).sV];       % lists all ventral boundaries
>> v = v';                 % transposition
>> d = [soln(:).sD];       % repeat for dorsal boundaries
>> d = d';
>> boxplot([v,d]);         % produce side-by-side boxplots
```

**Output 7:**




The two boxplots show the locations of the ventral (left, 1) and dorsal (right, 2) borders of the *vnd* gene expression domain.

## S2. Detailed supplementary methods

### S2.1. SETUP

All computational analysis was implemented in Matlab (R2010a or higher, Mathworks). Source code is organized in several folders. Before running individual functions, these folders should be placed in Matlab search path. Matlab provides a dialog to do so. Select

```
File | Set Path…
```

to open the path dialog. Then click

```
Add with Subfolders…
```

button to ensure all subfolders are added to the Matlab search path. Point to the local folder where the package downloaded and unpacked. The list of folders should reflect newly added folders. Save the search path and close the dialog to finish setup.

### S2.2. RUNNING ANALYSIS

The main entry point for the computational analysis is the `analyze_xs()` function. This function calls several other functions within the main analysis loop to perform individual tasks during analysis. Major tasks and related functions are detailed in the following subsections. Several bookkeeping steps before and after the main loop are as follows.

*S2.2.1. Preparation tasks:*

If not provided, optional arguments to `analyze_xs()` are assigned default values as follows:

> `yesplot`: whether or not to plot the image and the boundary. Default value is `false`.
> `ring_width`: width of ring in microns. Default value = 18.36.
> `stage`: nuclear cycle. Default value is 14.
> `nt`: number of radial bins. Default value is 60.

`analyze_xs()` calls `lsminfo()` to read metadata included in the LSM file and checks for the number of channels. Each channel is assigned a type according to the numerical identifiers given to `analyze_xs()`; see Section S1.5 of the Supplementary Manual for details.

The set of frames are read from the LSM file and scaled with respect to voxel sizes. Individual frames are then cropped to the region of interest containing the embryo cross-section data.

The background level (true black) is estimated for each channel independently, to be the most frequent intensity level (mode). This background level is subtracted from all slices in individual channels.

*S2.2.2. Bookkeeping tasks:*

When the main analysis loop completes, analysis results are populated in a structure (see Section S1.8.1 of Supplementary Manual).

This structure is also stored on disk in a Matlab data file, which is named as the original file appended with "`_data.mat`".

*S2.2.3. Output:*

The main function, `analyze_xs()`, returns a structure where individual fields store all information about input data and analysis results. Relevant fields of this structure may be updated by subsequent analysis operations. These fields are as follows:

`filename`: name of the input file
`channels`: the number code identifying measurement type in each channel
`channeltypes`: the type of each measurement in each channel
`channelnames`: description of what is in each channel
`genotype`: genotype of embryo
`metadata`: a structure containing metadata (see below)
`H`: the number of y-pixels
`W`: the number of x-pixels
`D`: the number of z-slices
`Lbar`: the average size of embryo (half-circumference) in microns
`s_mid`: 1-by-n_channels vector of where the midline is predicted to be. In most cases, the midline is the ventral midline, but in some (such as `pMad`), it would be the dorsal midline. Each channel gets an opportunity to predict where the midline is.
`s`: pseudo arclength, generated as `linspace(-1,1,301)'`
`t`: 301-by-n_channels array of the smoothened data as you go around the periphery of the embryo. Averaged in the z-direction.
`S`: the pseudo arclength coordinates of each nucleus. The z-direction is ignored.
`R`: the intensity of nuclear protein(s) for each nucleus. Same length as `S`, but has `np` columns, where `np` is the number of nuclear proteins. The identity of the nuclear protein in each column is in the same order as specified in `channelnames` field.

`Std_R`: the standard deviation of each measurement in `R`.
`Intron`: similar to `R`, but the intronic probe intensity for each nucleus.
`A,B,M,mu,sig`: parameters of Gaussian fits for the nuclear proteins
`dA,dB,dM,dmu,dsig`: length of 68% error-bars on above parameters
`gof`: r-square goodness-of-fit value for the Gaussian-fits
`sV,sD,w`: the ventral border, dorsal border, and widths of gene expression patterns in the mRNA or intronic probes. Each of these will be a 1-by-n_probe vector, where `n_probe` is the number of either mRNA or intronic probe channels. The identity of the gene in each column is in the same order as specified in `channelnames`.
`dsV,dsD,dw`: length of 68% error-bars on above parameters
`gof_gene`: the r-square goodness-of-fit value for the gene expression fits

The metadata field contains a structure with the following fields. These values collectively describe the data. The user is not expected to interact with metadata.

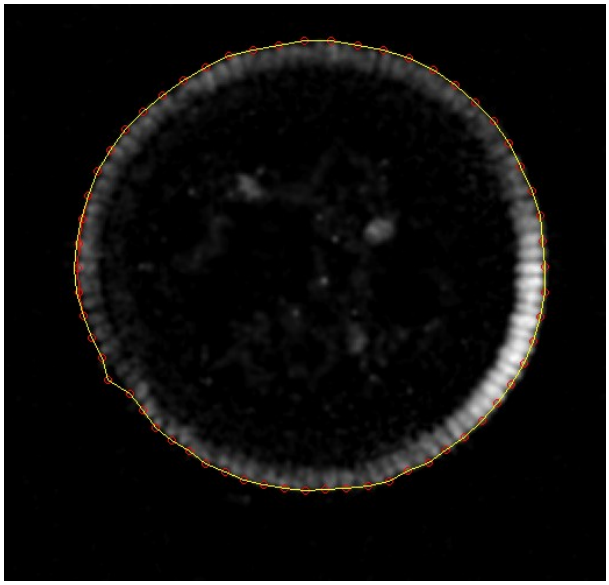
`lsminf1`: image metadata transferred from `lsmRead()`
`lsminf2`: image metadata transferred from `lsminfo()`
`scalings`: 1x3 vector of the x,y,z scalings in microns per pixel
`rho`: the nearest integer to the ratio of z scaling to xy scaling
`Yhatmax`: the distance into the embryo (in pixels) the nuclear layer is taken to be
`nt`: number of points in theta where the embryo periphery is evaluated
`bg`: the background levels of each of the channels
`std_bg`: the standard deviations of the background levels
`w`: the average arclength, in pixels, of the embryo periphery
`arc`: the arclength of each slice, in pixels
`L`: the arclength of each slice in microns of the embryo's half-periphery
`Xp,Yp`: cell array variables that contain the x and y coordinates of the periphery of the embryo. Each element of these cell arrays corresponds to a z-slice
`T,Raw`: the intensity of the mRNA channel as you go around the periphery of the embryo in quadrilaterals equally-spaced in periphery pseudo arclength. Each is a D-by-1 cell array. Each element in the cell arrays is a 301-by-n_channels array, where `n_channels` is the number of non-"N/A" channels.
`X,Y,S`: cell arrays containing the x,y and pseudoarclength coordinates of each nucleus. Each element of these cell arrays corresponds to a z-slice.
`Nuc,Std_nuc,Nuc_protein,Std_nuc_protein,Intron,Std_intron`: cell array variables that contain the intensity values and standard deviations of nuclei, nuclear proteins, and intronic probe intensity, for each nucleus. Each element of these cell arrays corresponds to a z-slice.
`cint,cint68`: the 95% and 68% confidence intervals on the gaussian fits to the nuclear proteins: [A B M mu sig]
`B_gene`: the background of gene expression fits.
`genes`: the metadata for the gene fits.
`introns`: the metadata for the intronic probe fits.

## S2.3. FINDING THE OUTER EDGE OF EMBRYO

The main steps of finding the outer edge of the embryo cross-section are as follows:
Image is smoothed with a Gaussian kernel. The cross-section is divided into 6-degree slices. Average intensity in each slice is found as a function of distance from the center of the cross-section. The radius where the intensity drops to 25% of maximum intensity marks the outer edge of the embryo. Finding the outer edge is implemented in `borderFinder()` function. This function takes an image, along with the following optional parameters:

`h`: height of intensity cutoff. Default value is 0.25.
`yesplot`: whether or not to plot the image and the boundary. Default value is `false`.
`nt`: choice for number of bins in theta. This value will determine the length of output arrays `xp` and `yp`. Default value is 60.

If `yesplot` is set to true, `borderFinder()` will produce a plot similar to the following:



The `borderFinder()` function returns the peripheral coordinates of the embryo cross-section in `xp` and `yp` arrays.

## S2.4. FINDING DOMAINS OF GENE EXPRESSION IN AN EMBRYO CROSS-SECTION

Once the outer edge of the cross-section was found, domains of gene expression are computed in the `domainMeas()` function as follows. Fluorescence intensity levels,

corresponding to gene expressions, are computed as a function of fractions of the total circumference of the embryo cross-section. Fractions of the circumference are approximated as trapezoidal regions in the image plane. `domainMeas()` function takes the fluorescence image and the coordinates of the outer edge. Optionally, the distance (in pixels) from the outer edge towards the center of the cross-section, i.e., the thickness between the inner and outer edge, can be provided as a parameter. If not provided, this value defaults to 30 pixels.

The cross-section is sampled at 300 points around its circumference. Then, gene expression levels are evaluated at each of these points for each channel. Intensity levels are smoothed by a moving average. The results are returned in the following format:

> `t` : (number of points) by (number of channels) array of smoothed fluorescent intensity levels
> `raw` : same as `t`, but the raw, non-smoothed data
> `s`: arc-length of each region, going from -1 to +1 divided into 300 points

## S2.5. SEGMENTING NUCLEI

If a nuclear channel is present, positions of each nuclei are segmented by the `find_nuclei()` function as follows. Image is first prepared by subtracting the background, approximated by the rolling ball algorithm with a 20-pixel disk, and smoothing with a Gaussian filter. Then, the cross-section image is "unrolled" into a linear representation of nuclei by the function `unroll2()`. The outer edge points, computed by `borderFinder()` function, are used as the landmarks for unrolling. First a trapezoidal region between the inner and outer edges of the cross-section is extracted at each point. Then, the trapezoidal regions are applied to a projective transformation based on their four vertices to covert them into full rectangles, effectively locally transforming the coordinate system. The resulting strip of nuclei are then returned as an image to the segmenting function, `find_nuclei()`.

The segmentation uses a coarse estimate of nuclear positions, which is calculated by the morphological opening operation. Then, a watershed is computed to determine trough and peak locations of each nuclei. Within each computed region between the troughs and peaks, a local threshold is applied to determine the nuclear position. Finally, a morphological opening of the image cleans possible spurious pixels. Nuclei centroids and pixels are extracted from the one-dimensional representation, which are subsequently used for reconstructing the result cross-section.

The function `find_nuclei()` takes the following input arguments:

`I`: image of the nuclei channel
`xp, yp`: points of embryo periphery from `borderFinder()`
`scalings`: row vector representing microns per pixel in x, y, z dimensions
`Yhatmax`: (optional) number of pixels between the inner and outer boundary of the embryo cross-section. If not provided, the default value is computed by dividing 18.36 microns by the scale factor (in microns per pixel) in x-direction.
`yesplot`: whether or not to plot the image and the boundary. Default value is `false`
`stage`: nuclear cycle between 10-14. Default value is 14.

`find_nuclei()` returns the results as:

`nucstats`: structure containing the nuclei statistics
`xnuc, ynuc`: centroid locations of each nucleus
`snuc`: the pseudo-arclength locations of each nucleus
`w`: the length of the unrolled image. Roughly equivalent to the pseudo-perimeter of the embryo in pixels
`mask`: label image of nuclei where each nucleus is identified by an integer

## S2.6. CALCULATING NUCLEAR INTENSITY LEVELS

The function `nuclearintensity()` takes only nuclear channels and nuclei statistics from `find_nuclei()` and computes the mean and standard error of the mean for each nucleus. The output is as follows:

`Y`: (number of nuclei) by (number of nuclear protein channels) array of mean intensities
`stdY`: (number of nuclei) by (number of nuclear protein channels) array of standard error

## S2.7. CALCULATING INTENSITY LEVELS OF INTRONIC PROBES

The function `intronicintensity()` takes only channels containing intronic probes and nuclei statistics from `find_nuclei()`. The nuclear dot intensity is taken to be the median intensity of a neighborhood, centered around the max intensity pixel and padded by 5-pixels in both directions. The output is as follows:

`Y`: (number of nuclei) by (number of channel) array of mean intensities
`stdY`: (number of nuclei) by (number of channel) array of standard error

## S2.8. ESTIMATING DORSAL NUCLEAR GRADIENT

The function `fit_gaussian()` estimates the parameters of a Gaussian PDF to the dorsal gradient data extracted by `analyze_xs()`. `fit_gaussian()` accepts the structure returned by the `analyze_xs()` as its input. The intensity profile of nuclear proteins is approximated as a Gaussian distribution. The closed form of the exact distribution used in fitting is given by:

$$f(x) = A * e^{\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)} + B + M|x|$$

and the non-linear least squares fit is used for parameter estimation.

The output format is similar to the input format, with the exception that the values of several fields are populated. These fields are as follows:

`data.s_mid` : row vector of length (number of total channels). Each channel (except for nuclei and N/A channels) has an independent estimate of the ventral midline. The value of the midline estmate from the nuclear protein channel(s) will be placed in the appropriate element of this vector

`data.nucprotein_names` : string identifying which nuclear protein (i.e., Dorsal, Sna, etc)

`data.A` : gradient amplitude, row vector of length (number of nuclear protein channels)

`data.B` : basal levels, row vector of length (number of nuclear protein channels)

`data.M` : slope of gradient tail, row vector of length (number of nuclear protein channels)

`data.mu` : the midline associated with the nuclear protein channel(s), row vector of length (number of nuclear protein channels)

`data.sig` : spatial extent of the gradient, row vector of length (number of nuclear protein channels)

`data.dA,.dB,.dM,.dmu,.dsig` : error bars of respective estimates above, row vectors of length (number of nuclear protein channels)

`data.gof` : $R^2$ goodness of fit for each nuclear channel, row vector of length (number of nuclear protein channels)

`data.metadata` : this field is now updated with the following metadata fields

`data.metadata.cint` : cell array of size one by (number of nuclear protein channels) Each element in the cell array is a two by five array containing the 95% confidence intervals on the five parameters (first row lower bound, second row upper bound) in the order A,B,M,mu,sig.

`data.metadata.cint68` : Same as above but for the 68% confidence intervals.

## S2.9. ESTIMATING NON-NUCLEAR GENES WITH CANONICAL PEAKS

The function `fit_peaks()` takes intensity values around the periphery of the embryo and fits canonical gene expression patterns for embryos in nuclear cycle 14. This is also done for intronic probe intensity values, which are first smoothed using the function `smooth_intron()`. In either case, the data are first background-subtracted before finding the ventral midline. Finding ventral midline is handled by the `find_midline()` function.

This function, including its ability to call `find_midline()` to find the ventral midline, requires the user to supply a channel name that is consistent with a known list of genes, which is hard-encoded into the beginning of `fit_peaks`. This is because the accuracy and/or success in finding peaks of gene expression (and the ventral midline) is dependent on the program knowing what kinds of peaks to expect. For example, if a channel name is '`vnd`', then the program knows there will be two peaks of gene expression, symmetric about the ventral midline, and each roughly 25% of the way from the ventral midline to the dorsal midline. When to strong peaks of signal are found in a rough search of the image, then the program knows to place the ventral midline directly in between the two peaks. Then, the fitting procedure knows where to begin looking to make a quantitative fit of the canonical profile to the measured profiles.

The main input for the `fit_peaks()` function is the same as in `fit_gaussian()`, i.e., the output structure from `analyze_xs()`. In addition, the following optional arguments are supported:

`yesplot` : whether or not to plot the image and the boundary. Default value is `false`.
`bkgrndthresh` : fraction of max height where background begins, default is 0.15.
`peakthresh` : fraction of max height of individual peaks end, default is 0.1.

The output of `fit_peaks()` is the same structure, except that the values of several fields are populated and updated as follows:

`data.s_mid` : row vector of length (number of total channels). Each channel (except for nuclei and N/A channels) has an independent estimate of the ventral midline. The value of the midline estmate from the mRNA and/or intron channels will be placed in the appropriate element of this vector
`data.gene_names` : string identifying genes
`data.sV` : location of ventral border of genes, in the order given in the gene_names field.
`data.sD` : location of dorsal border of genes, in the order given in the gene_names field.
`data.w` : location of width of genes, in the order given in the gene_names field.

`data.dsV,.dsD,.dw` : error bars of respective estimates above
`gof_gene` : goodness of fit in $R^2$ sense for each gene
`data.metadata` : this field is now updated with the following subfield
`data.metadata.genes` : contains further details on intermediate values used in fitting, such as $\delta$, $x_0$, $\alpha$, and $\beta$, as well as the 95% and 68% confidence intervals on these parameters.

---

**S3. Supplementary MATLAB scripts**

*MATLAB scripts and examples are available for download at journal website.*

APPENDICES

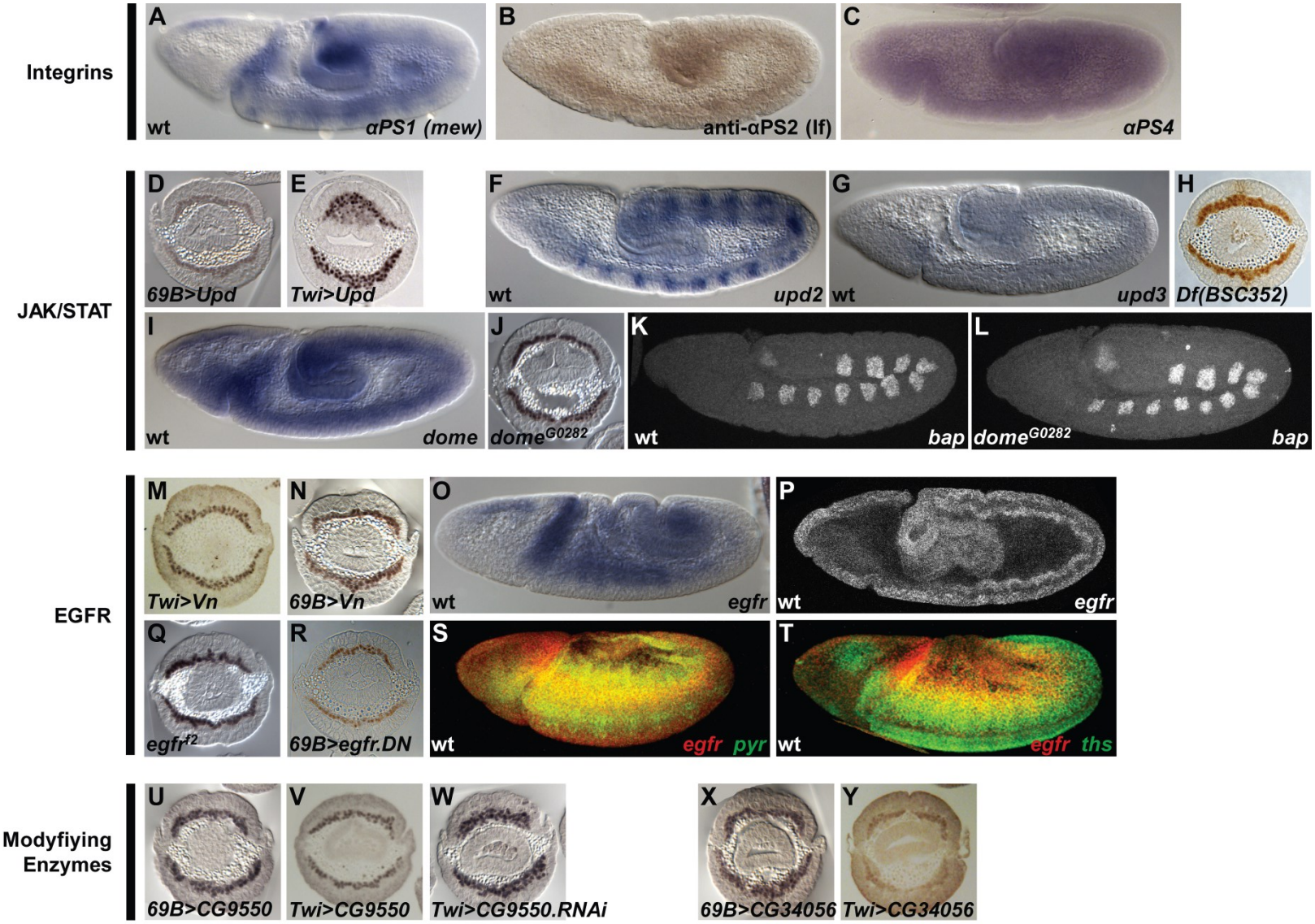
# C. Supplementary Materials for Chapter V

# SUPPLEMENTARY INVENTORY

**Supplementary Figure 1:** Expressions and mutant phenotypes of genes identified in screen.

**Supplementary Figure 2:** Mesoderm phenotypes when overexpressing or reducing additional HSPGs and CSPG.

**Supplementary Table 1:** Ectopic expression of twenty-four genes conferred lethality.


Integrins
A
wt
αPS1 (mew)
B
anti-αPS2 (lf)
C
αPS4
JAK/STAT
D
69B>Upd
E
Twi>Upd
F
wt
upd2
G
wt
upd3
H
Df(BSC352)
I
wt
dome
J
domeG0282
K
wt
bap
L
domeG0282
bap
EGFR
M
Twi>Vn
N
69B>Vn
O
wt
egfr
P
wt
egfr
Q
egfrf2
R
69B>egfr.DN
S
wt
egfr pyr
T
wt
egfr ths
Modyfiying Enzymes
U
69B>CG9550
V
Twi>CG9550
W
Twi>CG9550.RNAi
X
69B>CG34056
Y
Twi>CG34056

**Supplementary Figure 1. Expressions and mutant phenotypes of genes identified in screen.**

In situ hybridization was performed using riboprobes against the indicated genes. Lateral views of whole mount embryos are positioned with anterior facing left and dorsal side facing up. Cross-sectioned stage 10 embryos were stained with α-Twist to mark mesoderm cells.

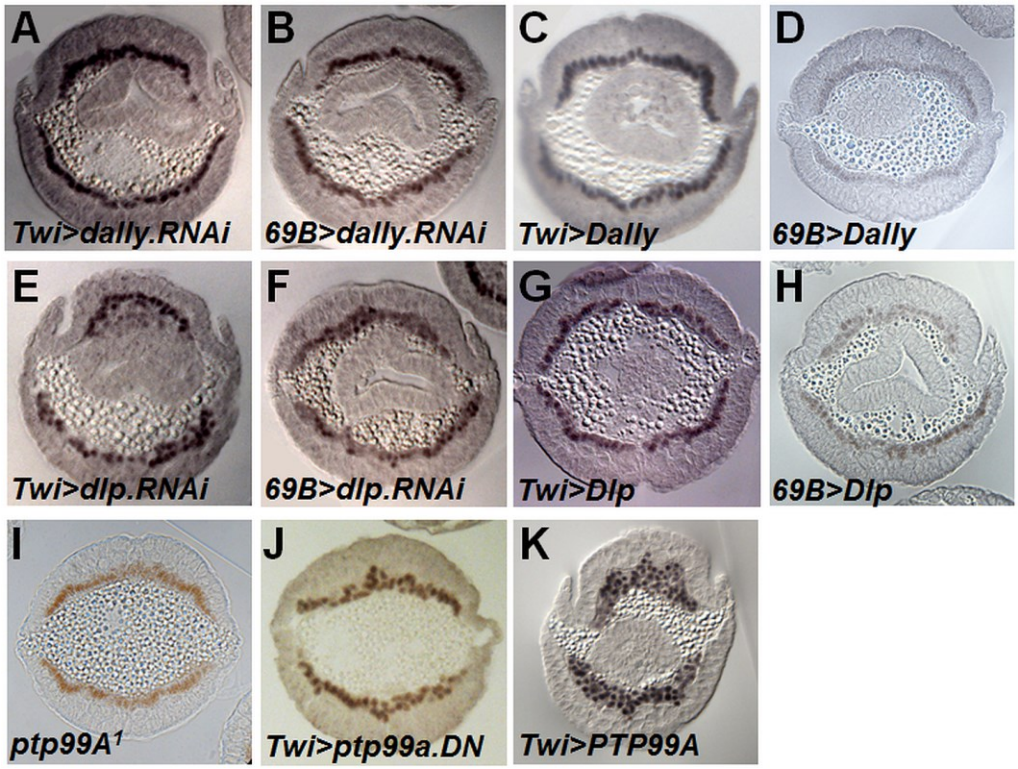
Integrins: Wildtype expression patterns for (A) αPS1, (B) αPS2 and (C) αPS4 show integrins are present at stage 8 embryos. α-PS2 (Inflated) was specifically found to be upregulated in the mesoderm (Y-K.Bae and A.S., unpub. obs.)

JAK/STAT: Cross-section of embryos overexpressing Upd in the (D) ectoderm and (E) mesoderm reveal multilayer phenotype. RNA expression of ligands (F) *upd2* and (G) *upd3* in wildtype embryos. (H) Cross-section of deficiency covering all three *upd* ligands has a mild spreading phenotype. (I) Wildtype expression of receptor *dome* shows upregulation in the mesoderm. Dome was also identified in a separate screen of mesoderm factors (Y-K.Bae and A.S., unpub. obs.). (J) Cross-section of *dome* mutant embryos have wildtype spreading. *bap* expression (Azpiazu and Frasch, 1993) in (K) wildtype is comparable to (L) *dome* mutant embryos, indicating normal mesoderm spreading.

EGFR: Cross-section of embryos overexpressing Vn in the (M) mesoderm and (N) ectoderm. Wildtype (O) stage 7 and (P) stage 10 embryos reveal *egfr* switch from ectodermal to mesodermal expression. Cross-section of embryos (Q) mutant for *egfr* or (R) overexpressing the dominant negative form of *egfr* in the ectoderm have relatively normal

spreading. Wildtype expression of *egfr* and FGF ligands (S) *pyr* and (T) *ths* show overlapping domains at stage 7, suggesting a possibility of EGFR affecting FGF ligands.

Modifying enzymes: Cross-section of embryos overexpressing CG9550 in the ectoderm have spreading defects (U), while overexpression in the mesoderm is normal (V). (W) Embryos removing *cg9550* by RNAi in the mesoderm result in a multilayer. Together, these date suggest a role for CG9550 in the mesoderm. Similarly, cross-sections of embryos overexpressing CG34056 in the ectoderm (X), but not mesoderm (Y), show spreading defects.




**Supplementary Figure 2. Mesoderm phenotypes when overexpressing or reducing additional HSPGs and CSPG.**

RNAi and overexpression mutant analysis for HSPGs Dally (A-D) and Dally-like (E-H), and CHSPG Ptp99a (I-K) revealed mild to no effects on mesoderm spreading. (J) DN refers to the dominant negative form of *ptp99a* in which the phosphatase domain was deleted.

| Gene ID | Name | | UAS | Lethality | Localization | Mutant allele(s) | Function |
|---|---|---|---|---|---|---|---|
| CG8084 | Anachronism | Ana | GS 9498 | Twi-Gal4 | secreted | Df(2R)BSC270 | growth factor activity |
| CG4531 | Argos | Aos | GS 12984 | 69B-Gal4 | secreted | Df(3L)BSC562 | antagonist of EGFR signalling |
| CG12086 | Cueball | Cue | EY 1263 | 69B-Gal4 | cell surface | $cue^{2}$ - hypomorph | LDLR class B repeats; EGF-like |
| CG15013 | Dusky-like | Dyl | GS 20894 | 69B-Gal4 | cell surface | Df(3L)BSC370 | zona pellucida domain |
| CG3722 | E-cadherin/ Shotgun | Shg | XP d01606 | 69B-Gal4 | cell surface | $shg^{2}$ - amorph | cadherin |
| CG32356 | Ecdysone-inducible gene E1 | ImpE1 | GS 11510 | 69B-Gal4 | cell surface | Df(3L)BSC815 | LDLR class A repeat |
| CG1106 | Gelsolin | Gel | GS 10156 | Twi-Gal4 | secreted | n/a | actin binding |
| CG32464 | l(3)82Fd/ Mustard | Mtd | GS 16948 | 69B-Gal4 | secreted | Df(3R)BSC139 | peptidoglycan-binding LysM domain |
| CG8434 | Lambik | Lbk | GS 17119 | Twi-Gal4 | cell surface | n/a | cell adhesion, Ig domain, LRR |
| CG7476 | Methuselah-like 7 | Mthl7 | GS 21256 | Twi-Gal4 | cell surface | Df(3L)BSC376 | GPCR |
| CG9342 | Microsomal triacylglycerol transfer protein | Mtp | XP d07488 | Twi-Gal4 | secreted | Df(2L)Exel7080 | lipid transport protein; triglyceride binding |
| CG2005 | Protein tyrosine phosphatase 99A | Ptp99A | EY 7423 | Twi-Gal4 | cell surface | $ptp99a^{1}$ - phosphatase domain delete; UAS-ptp99a.DN (K. Zinn) | Fibronectin type III domain |
| CG13194 | Pyramus | Pyr | GS 22603 | 69B-Gal4 | secreted | | FGF ligand |
| CG8095 | Scab | Scb | EP 2591 | Twi-Gal4 | cell surface | $scab^{2}$ | alpha-PS3 integrin |
| CG5661 | Semaphorin-5c | Sema-5c | EY 1704 | 69B-Gal4 | cell surface | Df(3L)BSC395 | Plexin; Sema domain |
| CG33950 | Terribly reduced optic lobes | Trol | GE 10067 | 69B-Gal4 | secreted | $trol^{G0211}$, FRT19A *111801; UAS-trol.RNAi §24549 | HSPG; Perlecan |
| CG6890 | Toll-8/Tollo | | XP d01565 | 69B-Gal4 | cell surface | Df(3L)BSC578 | LLR; Toll/interleukin-1 receptor homology (TIR) domain |
| CG5528 | Toll-9/Toll-like | | GS 51 | 69B-Gal4 | cell surface | Df(3L)BSC667 | LLR; Toll/interleukin-1 receptor homology (TIR) domain |
| CG9138 | Uninflatable | Uif | GS 11655 | 69B-Gal4 | cell surface | Df(2L)Exel7029 | LDLR class A repeat; EGF-like calium binding |
| CG5993 | Unpaired/ Outstretched | Upd/Os | G17133 | Twi-Gal4 & 69B-Gal4 | secreted | $upd^{4}$ - loss of fxn; Df(1)BSC352 | JAK/STAT ligand |
| CG10491 | Vein | Vn | GS 12044 | Twi-Gal4 & 69B-Gal4 | secreted | $vn^{C221}$ | EGFR ligand |
| CG5372 | αPS5 | | GS 12413 | 69B-Gal4 | cell surface | $αPS5^{[MI01533]}$ | integrin-related |
| CG34056 | CG34056 | | GS 11028 | 69B-Gal4 | cell surface | Df(3L)Exel6086 | beta-1,3-galactosyl-transferase activity |
| CG9550 | CG9550 | | GS 18034 | 69B-Gal4 | cell surface | Df(2L)BSC354; UAS-RNAi §100976 | chondroitin 6-sulfotransferase activity |
| ***Additional genes used in this study*** | | | | | | | |
| CG14226 | Domeless | Dome | | | cell surface | $dome^{G0282}$ - loss of fxn | JAK/STAT receptor |
| CG10079 | EGFR/ Torpedo | | | | cell surface | $egfr^{f2}$; UAS-egfr.DN #5364 | EGF Receptor |
| CG10497 | Syndecan | Sdc | | | cell surface | UAS-Sdc #8564; UAS-sdc.RNAi §13322 | HSPG |
| CG4974 | Dally | | | | cell surface | UAS-Dally #5397; UAS-dally.RNAi §14136 | HSPG |
| CG32146 | Dally-like | Dlp | | | cell surface | UAS-Dlp #9160; UAS-dlp.RNAi §10299 | HSPG |
| CG10275 | Kon-tiki/ Perdido | Kon/ Perd | | | cell surface | UAS-kon.RNAi §37283 | CSPG |

**# Bloomington Stock Center**
**§ VDRC Stock Center**
*** DGRC Kyoto Stock Center**

**Supplementary Table 1. Ectopic expression of twenty-four genes conferred lethality.**

Complete list of the twenty-four genes that resulted in lethality when overexpressed with Twi-Gal4 and/or 69B-Gal4. The genotypes used in this study and their predicted/known functions are also listed here. Genes in red indicate those with mesoderm spreading defects and/or relevant expression patterns and were further analyzed. Pyramus in blue has previously been well characterized. Additional genes that were examined in this study are noted as well.

# Bibliography

Alexandre, C., Baena-Lopez, A., Vincent, J.P., 2014. Patterning and growth control by membrane-tethered Wingless. Nature 505, 180-185.

Ali-Murthy, Z., Lott, S.E., Eisen, M.B., Kornberg, T.B., 2013. An essential role for zygotic expression in the pre-cellular Drosophila embryo. PLoS genetics 9, e1003428.

Allen, B.L., Rapraeger, A.C., 2003. Spatial and temporal expression of heparan sulfate in mouse development regulates FGF and FGF receptor assembly. The Journal of cell biology 163, 637-648.

Aman, A., Piotrowski, T., 2011. Cell-cell signaling interactions coordinate multiple cell behaviors that drive morphogenesis of the lateral line. Cell Adh Migr 5, 499-508.

Ay, A., Arnosti, D.N., 2011. Mathematical modeling of gene expression: a guide for the perplexed biologist. Critical reviews in biochemistry and molecular biology 46, 137-151.

Ay, A., Fakhouri, W.D., Chiu, C., Arnosti, D.N., 2008. Image processing and analysis for quantifying gene expression from early Drosophila embryos. Tissue engineering. Part A 14, 1517-1526.

Azpiazu, N., Frasch, M., 1993. tinman and bagpipe: two homeo box genes that determine cell fates in the dorsal mesoderm of Drosophila. Genes & development 7, 1325-1340.

Bae, Y.K., Trisnadi, N., Kadam, S., Stathopoulos, A., 2012. The role of FGF signaling in guiding coordinate movement of cell groups: guidance cue and cell adhesion regulator? Cell Adh Migr 6, 397-403.

Baeg, G.H., Selva, E.M., Goodman, R.M., Dasgupta, R., Perrimon, N., 2004. The Wingless morphogen gradient is established by the cooperative action of Frizzled and Heparan Sulfate Proteoglycan receptors. Dev Biol 276, 89-100.

Baum, B., Settleman, J., Quinlan, M.P., 2008. Transitions between epithelial and mesenchymal states in development and disease. Seminars in cell & developmental biology 19, 294-308.

Beiman, M., Shilo, B.Z., Volk, T., 1996. Heartless, a Drosophila FGF receptor homolog, is essential for cell migration and establishment of several mesodermal lineages. Genes Dev 10, 2993-3002.

Bergmann, S., Sandler, O., Sberro, H., Shnider, S., Schejter, E., Shilo, B.Z., Barkai, N., 2007. Pre-steady-state decoding of the Bicoid morphogen gradient. PLoS Biol 5, e46.

Bier, E., Bodmer, R., 2004. Drosophila, an emerging model for cardiac disease. Gene 342, 1-11.

Bischof, J., Maeda, R.K., Hediger, M., Karch, F., Basler, K., 2007. An optimized transgenesis system for Drosophila using germ-line-specific {varphi}C31 integrases. Proceedings of the National Academy of Sciences 104, 3312.

Boettiger, A.N., Levine, M., 2009. Synchronous and stochastic patterns of gene activation in the Drosophila embryo. Science 325, 471-473.

Bolouri, H., Davidson, E.H., 2003. Transcriptional regulatory cascades in development: initial rates, not steady state, determine network kinetics. Proc Natl Acad Sci U S A 100, 9371-9376.

Bothma, J.P., Levine, M., Boettiger, A., 2010. Morphogen gradients: limits to signaling or limits to measurement? Current biology : CB 20, R232-234.

Bothma, J.P., Magliocco, J., Levine, M., 2011. The snail repressor inhibits release, not elongation, of paused Pol II in the Drosophila embryo. Curr Biol 21, 1571-1577.

Bulgakova, N.A., Klapholz, B., Brown, N.H., 2012. Cell adhesion in Drosophila: versatility of cadherin and integrin complexes during development. Current opinion in cell biology 24, 702-712.

Buxton, P.G., Kostakopoulou, K., Brickell, P., Thorogood, P., Ferretti, P., 1997. Expression of the transcription factor slug correlates with growth of the limb bud and is regulated by FGF-4 and retinoic acid. The International journal of developmental biology 41, 559-568.

Cai, D., Chen, S.C., Prasad, M., He, L., Wang, X., Choesmel-Cadamuro, V., Sawyer, J.K., Danuser, G., Montell, D.J., 2014. Mechanical Feedback through E-Cadherin Promotes Direction Sensing during Collective Cell Migration. Cell 157, 1146-1159.

Campos-Ortega, J.A., Hartenstein, V., 1997. The embryonic development of Drosophila melanogaster, 2nd ed. Springer, Berlin ; New York.

Chung, K., Kim, Y., Kanodia, J.S., Gong, E., Shvartsman, S.Y., Lu, H., 2011. A microfluidic array for large-scale ordering and orientation of embryos. Nature methods 8, 171-176.

Ciruna, B., Rossant, J., 2001. FGF signaling regulates mesoderm cell fate specification and morphogenetic movement at the primitive streak. Developmental cell 1, 37-49.

Clark, I.B., Muha, V., Klingseisen, A., Leptin, M., Muller, H.A., 2011. Fibroblast growth factor signalling controls successive cell behaviours during mesoderm layer formation in Drosophila. Development 138, 2705-2715.

Cowden, J., Levine, M., 2002. The Snail repressor positions Notch signaling in the Drosophila embryo. Development 129, 1785-1793.

DeLotto, R., DeLotto, Y., Steward, R., Lippincott-Schwartz, J., 2007. Nucleocytoplasmic shuttling mediates the dynamic maintenance of nuclear Dorsal levels during Drosophila embryogenesis. Development 134, 4233-4241.

Dessaud, E., Ribes, V., Balaskas, N., Yang, L.L., Pierani, A., Kicheva, A., Novitch, B.G., Briscoe, J., Sasai, N., 2010. Dynamic assignment and maintenance of positional identity in the ventral neural tube by the morphogen sonic hedgehog. PLoS Biol 8, e1000382.

Dorfman, R., Shilo, B.Z., 2001. Biphasic activation of the BMP pathway patterns the Drosophila embryonic dorsal region. Development 128, 965-972.

Dragojlovic-Munther, M., Martinez-Agosto, J.A., 2013. Extracellular matrix-modulated Heartless signaling in Drosophila blood progenitors regulates their differentiation via a Ras/ETS/FOG pathway and target of rapamycin function. Dev Biol 384, 313-330.

Dunipace, L., Ozdemir, A., Stathopoulos, A., 2011. Complex interactions between cis-regulatory modules in native conformation are critical for Drosophila snail expression. Development 138, 4075-4084.

Dunipace, L., Saunders, A., Ashe, H.L., Stathopoulos, A., 2013. Autoregulatory feedback controls sequential action of cis-regulatory modules at the brinker locus. Developmental cell 26, 536-543.

Edgar, B.A., Kiehle, C.P., Schubiger, G., 1986. Cell cycle control by the nucleo-cytoplasmic ratio in early Drosophila development. Cell 44, 365-372.

Edgar, B.A., Sprenger, F., Duronio, R.J., Leopold, P., O'Farrell, P.H., 1994. Distinct molecular mechanism regulate cell cycle timing at successive stages of Drosophila embryogenesis. Genes Dev 8, 440-452.

Farach-Carson, M.C., Warren, C.R., Harrington, D.A., Carson, D.D., 2014. Border patrol: Insights into the unique role of perlecan/heparan sulfate proteoglycan 2 at cell and tissue borders. Matrix biology : journal of the International Society for Matrix Biology 34, 64-79.

Ferguson, E.L., Anderson, K.V., 1992. Localized enhancement and repression of the activity of the TGF-beta family member, decapentaplegic, is necessary for dorsal-ventral pattern formation in the Drosophila embryo. Development 114, 583-597.

Foe, V.E., 1989. Mitotic domains reveal early commitment of cells in Drosophila embryos. Development 107, 1-22.

Foe, V.E., Alberts, B.M., 1983. Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in Drosophila embryogenesis. Journal of cell science 61, 31-70.

Fowlkes, C.C., Hendriks, C.L., Keranen, S.V., Weber, G.H., Rubel, O., Huang, M.Y., Chatoor, S., DePace, A.H., Simirenko, L., Henriquez, C., Beaton, A., Weiszmann, R., Celniker, S., Hamann, B., Knowles, D.W., Biggin, M.D., Eisen, M.B., Malik, J., 2008. A quantitative spatiotemporal atlas of gene expression in the Drosophila blastoderm. Cell 133, 364-374.

Franzdottir, S.R., Engelen, D., Yuva-Aydemir, Y., Schmidt, I., Aho, A., Klambt, C., 2009. Switch in FGF signalling initiates glial differentiation in the Drosophila eye. Nature 460, 758-761.

Frasch, M., 1995. Induction of visceral and cardiac mesoderm by ectodermal Dpp in the early Drosophila embryo. Nature 374, 464-467.

Frasch, M., 1999. Intersecting signalling and transcriptional pathways in Drosophila heart specification. Seminars in cell & developmental biology 10, 61-71.

Friedl, P., Gilmour, D., 2009. Collective cell migration in morphogenesis, regeneration and cancer. Nat Rev Mol Cell Biol 10, 445-457.

Fung, T.K., Poon, R.Y., 2005. A roller coaster ride with the mitotic cyclins. Seminars in cell & developmental biology 16, 335-342.

Gabay, L., Seger, R., Shilo, B.Z., 1997a. In situ activation pattern of Drosophila EGF receptor pathway during development. Science 277, 1103-1106.

Gabay, L., Seger, R., Shilo, B.Z., 1997b. MAP kinase in situ activation atlas during Drosophila embryogenesis. Development 124, 3535-3541.

Garcia, H.G., Tikhonov, M., Lin, A., Gregor, T., 2013a. Quantitative imaging of transcription in living Drosophila embryos links polymerase activity to patterning. Curr Biol 23, 2140-2145.

Garcia, M., Nahmad, M., Reeves, G.T., Stathopoulos, A., 2013b. Size-dependent regulation of dorsal-ventral patterning in the early Drosophila embryo. Dev Biol 381, 286-299.

Garcia, M., Stathopoulos, A., 2011. Lateral gene expression in Drosophila early embryos is supported by Grainyhead-mediated activation and tiers of dorsally-localized repression. PLoS One 6, e29172.

Ghabrial, A.S., Krasnow, M.A., 2006. Social interactions among epithelial cells during tracheal branching morphogenesis. Nature 441, 746-749.

Gisselbrecht, S., Skeath, J.B., Doe, C.Q., Michelson, A.M., 1996. heartless encodes a fibroblast growth factor receptor (DFR1/DFGF-R2) involved in the directional migration of early mesodermal cells in the Drosophila embryo. Genes Dev 10, 3003-3017.

Goentoro, L.A., Yakoby, N., Goodhouse, J., Schupbach, T., Shvartsman, S.Y., 2006. Quantitative analysis of the GAL4/UAS system in Drosophila oogenesis. Genesis (New York, N.Y. : 2000) 44, 66-74.

Gregor, T., Bialek, W., de Ruyter van Steveninck, R.R., Tank, D.W., Wieschaus, E.F., 2005. Diffusion and scaling during early embryonic pattern formation. Proc Natl Acad Sci U S A 102, 18403-18407.

Gregor, T., Tank, D.W., Wieschaus, E.F., Bialek, W., 2007a. Probing the limits to positional information. Cell 130, 153-164.

Gregor, T., Wieschaus, E.F., McGregor, A.P., Bialek, W., Tank, D.W., 2007b. Stability and nuclear dynamics of the bicoid morphogen gradient. Cell 130, 141-152.

Grigorian, M., Liu, T., Banerjee, U., Hartenstein, V., 2013. The proteoglycan Trol controls the architecture of the extracellular matrix and balances proliferation and differentiation of blood progenitors in the Drosophila lymph gland. Dev Biol 384, 301-312.

Groth, A.C., Fish, M., Nusse, R., Calos, M.P., 2004. Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. Genetics 166, 1775-1782.

Gryzik, T., Muller, H.A., 2004. FGF8-like1 and FGF8-like2 encode putative ligands of the FGF receptor Htl and are required for mesoderm migration in the Drosophila gastrula. Curr Biol 14, 659-667.

Guo, Z., Wang, Z., 2009. The glypican Dally is required in the niche for the maintenance of germline stem cells and short-range BMP signaling in the Drosophila ovary. Development 136, 3627-3635.

Gurdon, J.B., Mitchell, A., Mahony, D., 1995. Direct and continuous assessment by cells of their position in a morphogen gradient. Nature 376, 520-521.

Hammerschmidt, M., Wedlich, D., 2008. Regulated adhesion as a driving force of gastrulation movements. Development 135, 3625-3641.

Harris, T.J., Sawyer, J.K., Peifer, M., 2009. How the cytoskeleton helps build the embryonic body plan: models of morphogenesis from Drosophila. Curr Top Dev Biol 89, 55-85.

Hatan, M., Shinder, V., Israeli, D., Schnorrer, F., Volk, T., 2011. The Drosophila blood brain barrier is maintained by GPCR-dependent dynamic actin structures. The Journal of cell biology 192, 307-319.

He, F., Wen, Y., Deng, J., Lin, X., Lu, L.J., Jiao, R., Ma, J., 2008. Probing intrinsic properties of a robust morphogen gradient in Drosophila. Developmental cell 15, 558-567.

Hirst, J., Carmichael, J., 2011. A potential role for the clathrin adaptor GGA in Drosophila spermatogenesis. BMC cell biology 12, 22.

Houchmandzadeh, B., Wieschaus, E., Leibler, S., 2002. Establishment of developmental precision and proportions in the early Drosophila embryo. Nature 415, 798-802.

Huang, A.M., Rusch, J., Levine, M., 1997. An anteroposterior Dorsal gradient in the Drosophila embryo. Genes Dev 11, 1963-1973.

Ismat, A., Schaub, C., Reim, I., Kirchner, K., Schultheis, D., Frasch, M., 2010. HLH54F is required for the specification and migration of longitudinal gut muscle founders from the caudal mesoderm of Drosophila. Development 137, 3107-3117.

Jaeger, J., Reinitz, J., 2006. On the dynamic nature of positional information. Bioessays 28, 1102-1111.

Jaeger, J., Surkova, S., Blagov, M., Janssens, H., Kosman, D., Kozlov, K.N., Manu, Myasnikova, E., Vanario-Alonso, C.E., Samsonova, M., Sharp, D.H., Reinitz, J., 2004. Dynamic control of positional information in the early Drosophila embryo. Nature 430, 368-371.

Jiang, J., Levine, M., 1993. Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. Cell 72, 741-752.

Johnson, A.N., Mokalled, M.H., Haden, T.N., Olson, E.N., 2011. JAK/Stat signaling regulates heart precursor diversification in Drosophila. Development 138, 4627-4638.

Kadam, S., Ghosh, S., Stathopoulos, A., 2012. Synchronous and symmetric migration of Drosophila caudal visceral mesoderm cells requires dual input by two FGF ligands. Development 139, 699-708.

Kadam, S., McMahon, A., Tzou, P., Stathopoulos, A., 2009. FGF ligands in Drosophila have distinct activities required to support cell migration and differentiation. Development 136, 739-747.

Kamimura, K., Fujise, M., Villa, F., Izumi, S., Habuchi, H., Kimata, K., Nakato, H., 2001. Drosophila heparan sulfate 6-O-sulfotransferase (dHS6ST) gene. Structure, expression, and function in the formation of the tracheal system. J Biol Chem 276, 17014-17021.

Kamimura, K., Ueno, K., Nakagawa, J., Hamada, R., Saitoe, M., Maeda, N., 2013. Perlecan regulates bidirectional Wnt signaling at the Drosophila neuromuscular junction. The Journal of cell biology 200, 219-233.

Kanodia, J.S., Rikhy, R., Kim, Y., Lund, V.K., DeLotto, R., Lippincott-Schwartz, J., Shvartsman, S.Y., 2009. Dynamics of the Dorsal morphogen gradient. Proc Natl Acad Sci U S A 106, 21707-21712.

Kim, S.H., Turnbull, J., Guimond, S., 2011. Extracellular matrix and cell signalling: the dynamic cooperation of integrin, proteoglycan and growth factor receptor. The Journal of endocrinology 209, 139-151.

Klingseisen, A., Clark, I.B., Gryzik, T., Muller, H.A., 2009. Differential and overlapping functions of two closely related Drosophila FGF8-like growth factors in mesoderm development. Development 136, 2393-2402.

Knox, J., Moyer, K., Yacoub, N., Soldaat, C., Komosa, M., Vassilieva, K., Wilk, R., Hu, J., Vazquez Paz Lde, L., Syed, Q., Krause, H.M., Georgescu, M., Jacobs, J.R., 2011. Syndecan contributes to heart cell specification and lumen formation during Drosophila cardiogenesis. Dev Biol 356, 279-290.

Kosman, D., Mizutani, C.M., Lemons, D., Cox, W.G., McGinnis, W., Bier, E., 2004. Multiplex detection of RNA expression in Drosophila embryos. Science 305, 846.

Kulkarni, M.M., 2011. Digital multiplexed gene expression analysis using the NanoString nCounter system. Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.] Chapter 25, Unit25B.10.

Kurusu, M., Cording, A., Taniguchi, M., Menon, K., Suzuki, E., Zinn, K., 2008. A screen of cell-surface molecules identifies leucine-rich repeat proteins as key mediators of synaptic target selection. Neuron 59, 972-985.

Kusch, T., Reuter, R., 1999. Functions for Drosophila brachyenteron and forkhead in mesoderm specification and cell signalling. Development 126, 3991-4003.

Kutejova, E., Briscoe, J., Kicheva, A., 2009. Temporal dynamics of patterning by morphogen gradients. Curr Opin Genet Dev 19, 315-322.

Kvon, E.Z., Kazmar, T., Stampfel, G., Yanez-Cuna, J.O., Pagani, M., Schernhuber, K., Dickson, B.J., Stark, A., 2014. Genome-scale functional characterization of Drosophila developmental enhancers in vivo. Nature.

Lecuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, T.R., Tomancak, P., Krause, H.M., 2007. Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. Cell 131, 174-187.

Lee, H.-H., Zaffran, S., Frasch, M., 2006. Development of the Larval Visceral Musculature, in: Sink, H. (Ed.), Muscle Development in Drosophila. Landes Bioscience, New York, N.Y., pp. 62-78.

Lee, L.A., Orr-Weaver, T.L., 2003. Regulation of cell cycles in Drosophila development: intrinsic and extrinsic cues. Annual review of genetics 37, 545-578.

Lek, M., Dias, J.M., Marklund, U., Uhde, C.W., Kurdija, S., Lei, Q., Sussel, L., Rubenstein, J.L., Matise, M.P., Arnold, H.H., Jessell, T.M., Ericson, J., 2010. A homeodomain feedback circuit underlies step-function interpretation of a Shh morphogen gradient during ventral neural patterning. Development 137, 4051-4060.

Liang, H.L., Nien, C.Y., Liu, H.Y., Metzstein, M.M., Kirov, N., Rushlow, C., 2008. The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature 456, 400-403.

Liberman, L.M., Reeves, G.T., Stathopoulos, A., 2009. Quantitative imaging of the Dorsal nuclear gradient reveals limitations to threshold-dependent patterning in Drosophila. Proceedings of the National Academy of Sciences of the United States of America 106, 22317-22322.

Liberman, L.M., Stathopoulos, A., 2009. Design flexibility in cis-regulatory control of gene expression: synthetic and comparative evidence. Dev Biol 327, 578-589.

Lin, X., 2004. Functions of heparan sulfate proteoglycans in cell signaling during development. Development 131, 6009-6021.

Lin, X., Buff, E.M., Perrimon, N., Michelson, A.M., 1999. Heparan sulfate proteoglycans are essential for FGF receptor signaling during Drosophila embryonic development. Development 126, 3715-3723.

Little, S.C., Tkacik, G., Kneeland, T.B., Wieschaus, E.F., Gregor, T., 2011. The formation of the Bicoid morphogen gradient requires protein movement from anteriorly localized mRNA. PLoS Biol 9, e1000596.

Lott, S.E., Villalta, J.E., Schroth, G.P., Luo, S., Tonkin, L.A., Eisen, M.B., 2011. Noncanonical compensation of zygotic X transcription in early Drosophila melanogaster development revealed through single-embryo RNA-seq. PLoS Biol 9, e1000590.

Lu, X., Drocco, J., Wieschaus, E.F., 2010. Cell cycle regulation via inter-nuclear communication during the early embryonic development of Drosophila melanogaster. Cell cycle (Georgetown, Tex.) 9, 2908-2910.

Lu, X., Li, J.M., Elemento, O., Tavazoie, S., Wieschaus, E.F., 2009. Coupling of zygotic transcription to mitotic control at the Drosophila mid-blastula transition. Development 136, 2101-2110.

Lucas, T., Ferraro, T., Roelens, B., De Las Heras Chanes, J., Walczak, A.M., Coppey, M., Dostatni, N., 2013. Live imaging of bicoid-dependent transcription in Drosophila embryos. Curr Biol 23, 2135-2139.

Lucchetta, E.M., Vincent, M.E., Ismagilov, R.F., 2008. A precise Bicoid gradient is nonessential during cycles 11-13 for precise patterning in the Drosophila blastoderm. PLoS One 3, e3651.

Luengo Hendriks, C.L., Keranen, S.V., Fowlkes, C.C., Simirenko, L., Weber, G.H., DePace, A.H., Henriquez, C., Kaszuba, D.W., Hamann, B., Eisen, M.B., Malik, J., Sudar, D., Biggin, M.D., Knowles, D.W., 2006. Three-dimensional morphology and gene expression in the Drosophila blastoderm at cellular resolution I: data acquisition pipeline. Genome biology 7, R123.

Mandal, L., Dumstrei, K., Hartenstein, V., 2004. Role of FGFR signaling in the morphogenesis of the Drosophila visceral musculature. Dev Dyn 231, 342-348.

Manu, Surkova, S., Spirov, A.V., Gursky, V.V., Janssens, H., Kim, A.R., Radulescu, O., Vanario-Alonso, C.E., Sharp, D.H., Samsonova, M., Reinitz, J., 2009. Canalization of gene expression in the Drosophila blastoderm by gap gene cross regulation. PLoS Biol 7, e1000049.

Mazumdar, A., Mazumdar, M., 2002. How one becomes many: blastoderm cellularization in Drosophila melanogaster. Bioessays 24, 1012-1022.

McCleland, M.L., Farrell, J.A., O'Farrell, P.H., 2009. Influence of cyclin type and dose on mitotic entry and progression in the early Drosophila embryo. The Journal of cell biology 184, 639-646.

McHale, P., Mizutani, C.M., Kosman, D., Mackay, D.L., Belu, M., Hermann, A., McGinnis, W., Bier, E., Hwa, T., 2011. Gene length may contribute to graded transcriptional responses in the Drosophila embryo. Dev Biol.

McMahon, A., Reeves, G.T., Supatto, W., Stathopoulos, A., 2010. Mesoderm migration in Drosophila is a multi-step process requiring FGF signaling and integrin activity. Development 137, 2167-2175.

McMahon, A., Supatto, W., Fraser, S.E., Stathopoulos, A., 2008. Dynamic analyses of Drosophila gastrulation provide insights into collective cell migration. Science 322, 1546-1550.

Montell, D.J., 2006. The social lives of migrating cells in Drosophila. Curr Opin Genet Dev 16, 374-383.

Moussian, B., Roth, S., 2005. Dorsoventral axis formation in the Drosophila embryo--shaping and transducing a morphogen gradient. Curr Biol 15, R887-899.

Mukherjee, T., Choi, I., Banerjee, U., 2012. Genetic analysis of fibroblast growth factor signaling in the Drosophila eye. G3 (Bethesda, Md.) 2, 23-28.

Murray, A.W., Kirschner, M.W., 1989. Cyclin synthesis drives the early embryonic cell cycle. Nature 339, 275-280.

Murray, M.J., Saint, R., 2007. Photoactivatable GFP resolves Drosophila mesoderm migration behaviour. Development 134, 3975-3983.

Nahmad, M., Stathopoulos, A., 2009. Dynamic interpretation of hedgehog signaling in the Drosophila wing disc. PLoS Biol 7, e1000202.

Nambu, J.R., Lewis, J.O., Wharton, K.A., Jr., Crews, S.T., 1991. The Drosophila single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development. Cell 67, 1157-1167.

O'Farrell, P.H., 2001. Triggering the all-or-nothing switch into mitosis. Trends in cell biology 11, 512-519.

Ochoa-Espinosa, A., Yu, D., Tsirigos, A., Struffi, P., Small, S., 2009. Anterior-posterior positional information in the absence of a strong Bicoid gradient. Proc Natl Acad Sci U S A 106, 3823-3828.

Oda, H., Tsukita, S., Takeichi, M., 1998. Dynamic behavior of the cadherin-based cell-cell adhesion system during Drosophila gastrulation. Dev Biol 203, 435-450.

Ornitz, D.M., 2000. FGFs, heparan sulfate and FGFRs: complex interactions essential for development. Bioessays 22, 108-112.

Otsu, N., 1979. A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man, and Cybernetics 9, 62-66.

Ozdemir, A., Fisher-Aylor, K.I., Pepke, S., Samanta, M., Dunipace, L., McCue, K., Zeng, L., Ogawa, N., Wold, B.J., Stathopoulos, A., 2011. High resolution mapping of Twist to DNA in Drosophila embryos: Efficient functional analysis and evolutionary conservation. Genome research 21, 566-577.

Park, Y., Rangel, C., Reynolds, M.M., Caldwell, M.C., Johns, M., Nayak, M., Welsh, C.J., McDermott, S., Datta, S., 2003. Drosophila perlecan modulates FGF and hedgehog signals to activate neural stem cell division. Dev Biol 253, 247-257.

Pellegrini, L., Burke, D.F., von Delft, F., Mulloy, B., Blundell, T.L., 2000. Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin. Nature 407, 1029-1034.

Perkins, T.J., Jaeger, J., Reinitz, J., Glass, L., 2006. Reverse engineering the gap gene network of Drosophila melanogaster. PLoS Comput Biol 2, e51.

Plotnikov, A.N., Hubbard, S.R., Schlessinger, J., Mohammadi, M., 2000. Crystal structures of two FGF-FGFR complexes reveal the determinants of ligand-receptor specificity. Cell 101, 413-424.

Porcher, A., Dostatni, N., 2010. The bicoid morphogen system. Curr Biol 20, R249-254.

Ramos, C., Becerril, C., Montano, M., Garcia-De-Alba, C., Ramirez, R., Checa, M., Pardo, A., Selman, M., 2010. FGF-1 reverts epithelial-mesenchymal transition induced by TGF-{beta}1 through MAPK/ERK kinase pathway. American journal of physiology. Lung cellular and molecular physiology 299, L222-231.

Reeves, G.T., Stathopoulos, A., 2009. Graded Dorsal and Differential Gene Regulation in the *Drosophila* Embryo, in: Briscoe, J., Lawrence, P., Vincent, J.-P. (Eds.), Perspectives on Generation and Interpretation of Morphogen Gradients. Cold Spring Harbor Laboratory Press.

Reeves, G.T., Trisnadi, N., Truong, T.V., Nahmad, M., Katz, S., Stathopoulos, A., 2012. Dorsal-ventral gene expression in the Drosophila embryo reflects the dynamics and precision of the dorsal nuclear gradient. Developmental cell 22, 544-557.

Reim, I., Hollfelder, D., Ismat, A., Frasch, M., 2012. The FGF8-related signals Pyramus and Thisbe promote pathfinding, substrate adhesion, and survival of migrating longitudinal gut muscle founder cells. Dev Biol 368, 28-43.

Rembold, M., Ciglar, L., Yanez-Cuna, J.O., Zinzen, R.P., Girardot, C., Jain, A., Welte, M.A., Stark, A., Leptin, M., Furlong, E.E., 2014. A conserved role for Snail as a potentiator of active transcription. Genes Dev 28, 167-181.

Ridley, A.J., Schwartz, M.A., Burridge, K., Firtel, R.A., Ginsberg, M.H., Borisy, G., Parsons, J.T., Horwitz, A.R., 2003. Cell migration: integrating signals from front to back. Science 302, 1704-1709.

Rorth, P., 2011. Whence directionality: guidance mechanisms in solitary and collective cell migration. Developmental cell 20, 9-18.

Rothe, M., Pehl, M., Taubert, H., Jackle, H., 1992. Loss of gene function through rapid mitotic cycles in the Drosophila embryo. Nature 359, 156-159.

Roy, S., Huang, H., Liu, S., Kornberg, T.B., 2014. Cytoneme-mediated contact-dependent transport of the Drosophila decapentaplegic signaling protein. Science 343, 1244624.

Rushlow, C., Colosimo, P.F., Lin, M.C., Xu, M., Kirov, N., 2001. Transcriptional regulation of the Drosophila gene zen by competing Smad and Brinker inputs. Genes & development 15, 340-351.

Sandmann, T., Girardot, C., Brehme, M., Tongprasit, W., Stolc, V., Furlong, E.E., 2007. A core transcriptional network for early mesoderm development in Drosophila melanogaster. Genes Dev 21, 436-449.

Sarrazin, S., Lamanna, W.C., Esko, J.D., 2011. Heparan sulfate proteoglycans. Cold Spring Harbor perspectives in biology 3.

Schafer, G., Narasimha, M., Vogelsang, E., Leptin, M., 2014. Cadherin switching during the formation and differentiation of the Drosophila mesoderm - implications for epithelial-to-mesenchymal transitions. Journal of cell science 127, 1511-1522.

Schneider, M., Baumgartner, S., 2008. Differential expression of Dystroglycan-spliceforms with and without the mucin-like domain during Drosophila embryogenesis. Fly 2, 29-35.

Shermoen, A.W., McCleland, M.L., O'Farrell, P.H., 2010. Developmental control of late replication and S phase length. Curr Biol 20, 2067-2077.

Shermoen, A.W., O'Farrell, P.H., 1991. Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. Cell 67, 303-310.

Shilo, B.Z., 2005. Regulating the dynamics of EGF receptor signaling in space and time. Development 132, 4017-4027.

Shimokawa, K., Kimura-Yoshida, C., Nagai, N., Mukai, K., Matsubara, K., Watanabe, H., Matsuda, Y., Mochida, K., Matsuo, I., 2011. Cell surface heparan sulfate chains regulate local reception of FGF signaling in the mouse embryo. Developmental cell 21, 257-272.

Shishido, E., Higashijima, S., Emori, Y., Saigo, K., 1993. Two FGF-receptor homologues of Drosophila: one is expressed in mesodermal primordium in early embryos. Development 117, 751-761.

Sibon, O.C., Stevenson, V.A., Theurkauf, W.E., 1997. DNA-replication checkpoint control at the Drosophila midblastula transition. Nature 388, 93-97.

Smart, A.D., Course, M.M., Rawson, J., Selleck, S., Van Vactor, D., Johnson, K.G., 2011. Heparan sulfate proteoglycan specificity during axon pathway formation in the Drosophila embryo. Developmental neurobiology 71, 608-618.

Smith, S.M., West, L.A., Govindraj, P., Zhang, X., Ornitz, D.M., Hassell, J.R., 2007. Heparan and chondroitin sulfate on growth plate perlecan mediate binding and delivery of FGF-2 to FGF receptors. Matrix biology : journal of the International Society for Matrix Biology 26, 175-184.

Sokac, A.M., Wieschaus, E., 2008. Local actin-dependent endocytosis is zygotically controlled to initiate Drosophila cellularization. Developmental cell 14, 775-786.

Solnica-Krezel, L., Sepich, D.S., 2012. Gastrulation: making and shaping germ layers. Annual review of cell and developmental biology 28, 687-717.

Song, Y., Ori-McKenney, K.M., Zheng, Y., Han, C., Jan, L.Y., Jan, Y.N., 2012. Regeneration of Drosophila sensory neuron axons and dendrites is regulated by the Akt pathway involving Pten and microRNA bantam. Genes Dev 26, 1612-1625.

Spitz, F., Furlong, E.E., 2012. Transcription factors: from enhancer binding to developmental control. Nature reviews. Genetics 13, 613-626.

Stathopoulos, A., Levine, M., 2005. Genomic regulatory networks and animal development. Developmental cell 9, 449-462.

Stathopoulos, A., Tam, B., Ronshaugen, M., Frasch, M., Levine, M., 2004. pyramus and thisbe: FGF genes that pattern the mesoderm of Drosophila embryos. Genes Dev 18, 687-699.

Sung, H.W., Spangenberg, S., Vogt, N., Grosshans, J., 2013. Number of nuclear divisions in the Drosophila blastoderm controlled by onset of zygotic transcription. Curr Biol 23, 133-138.

Supatto, W., McMahon, A., Fraser, S.E., Stathopoulos, A., 2009. Quantitative imaging of collective cell migration during Drosophila gastrulation: multiphoton microscopy and computational analysis. Nat Protoc 4, 1397-1412.

Surkova, S.Y., Myasnikova, E.M., Kozlov, K.N., Samsonova, A.A., Reinitz, J., Samsonova, M.G., 2008. Methods for Acquisition of Quantitative Data from Confocal Images of Gene Expression in situ. Cell and tissue biology 2, 200-215.

Tadros, W., Lipshitz, H.D., 2009. The maternal-to-zygotic transition: a play in two acts. Development 136, 3033-3042.

Tanaka, K.M., Takahashi, A., Fuse, N., Takano-Shimizu-Kouno, T., 2014. A Novel Cell Death Gene Acts to Repair Patterning Defects in Drosophila melanogaster. Genetics 197, 739-742.

Tepass, U., Hartenstein, V., 1994. The development of cellular junctions in the Drosophila embryo. Dev Biol 161, 563-596.

Theveneau, E., Marchant, L., Kuriyama, S., Gull, M., Moepps, B., Parsons, M., Mayor, R., 2010. Collective chemotaxis requires contact-dependent cell polarity. Developmental cell 19, 39-53.

Thummel, C.S., Burtis, K.C., Hogness, D.S., 1990. Spatial and temporal patterns of E74 transcription during Drosophila development. Cell 61, 101-111.

Tostevin, F., ten Wolde, P.R., Howard, M., 2007. Fundamental limits to position determination by concentration gradients. PLoS Comput Biol 3, e78.

Trisnadi, N., Altinok, A., Stathopoulos, A., Reeves, G.T., 2013. Image analysis and empirical modeling of gene and protein expression. Methods (San Diego, Calif.) 62, 68-78.

Truong, T.V., Supatto, W., Koos, D.S., Choi, J.M., Fraser, S.E., 2011. Deep and fast live imaging with two-photon scanned light-sheet microscopy. Nature methods.

Tulin, S., Stathopoulos, A., 2010a. Analysis of Thisbe and Pyramus functional domains reveals evidence for cleavage of Drosophila FGFs. BMC Dev Biol 10, 83.

Tulin, S., Stathopoulos, A., 2010b. Extending the family table: Insights from beyond vertebrates into the regulation of embryonic development by FGFs. Birth Defects Res C Embryo Today 90, 214-227.

Umulis, D.M., Shimmi, O., O'Connor, M.B., Othmer, H.G., 2010. Organism-scale modeling of early Drosophila patterning via bone morphogenetic proteins. Developmental cell 18, 260-274.

Urbano, J.M., Dominguez-Gimenez, P., Estrada, B., Martin-Bermudo, M.D., 2011. PS Integrins and Laminins: Key Regulators of Cell Migration during Drosophila Embryogenesis. PLoS One 6, e23893.

Venken, K.J., He, Y., Hoskins, R.A., Bellen, H.J., 2006. P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster. Science 314, 1747-1751.

Von Ohlen, T.L., Harvey, C., Panda, M., 2007. Identification of an upstream regulatory element reveals a novel requirement for Ind activity in maintaining ind expression. Mech Dev 124, 230-236.

Wang, X., He, L., Wu, Y.I., Hahn, K.M., Montell, D.J., 2010. Light-mediated activation reveals a key role for Rac in collective guidance of cell movement in vivo. Nature cell biology 12, 591-597.

Warming, S., Costantino, N., Court, D.L., Jenkins, N.A., Copeland, N.G., 2005. Simple and highly efficient BAC recombineering using galK selection. Nucleic Acids Res 33, e36.

Welch, B.L., 1947. THE GENERALIZATION OF 'STUDENT'S' PROBLEM WHEN SEVERAL DIFFERENT POPULATION VARIANCES ARE INVOLVED. Biometrika 34, 28-35.

Wilson, R., Vogelsang, E., Leptin, M., 2005. FGF signalling and the mechanism of mesoderm spreading in Drosophila embryos. Development 132, 491-501.

Winklbauer, R., Muller, H.A., 2011. Mesoderm layer formation in Xenopus and Drosophila gastrulation. Physical biology 8, 045001.

Wolpert, L., 1969. Positional information and the spatial pattern of cellular differentiation. Journal of theoretical biology 25, 1-47.

Wu, W., Tholozan, F.M., Goldberg, M.W., Bowen, L., Wu, J., Quinlan, R.A., 2014. A gradient of matrix-bound FGF-2 and perlecan is available to lens epithelial cells. Experimental eye research 120, 10-14.

Yuan, K., Shermoen, A.W., O'Farrell, P.H., 2014. Illuminating DNA replication during Drosophila development using TALE-lights. Curr Biol 24, R144-145.

Zeitlinger, J., Zinzen, R.P., Stark, A., Kellis, M., Zhang, H., Young, R.A., Levine, M., 2007. Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the Drosophila embryo. Genes Dev 21, 385-390.

Zhang, X., Ibrahimi, O.A., Olsen, S.K., Umemori, H., Mohammadi, M., Ornitz, D.M., 2006. Receptor specificity of the fibroblast growth factor family. The complete mammalian FGF family. J Biol Chem 281, 15694-15700.

Zinzen, R., Senger, K., Levine, M., Papatsenko, D., 2006. Computational Models for Neurogenic Gene Expression in the Drosophila Embryo. Current Biology 16, 1358-1365.